ISLET NEOGENESIS ASSOCIATED PROTEIN-RELATED PROTEIN: FROM GENE TO FOLDED PROTEIN

A Dissertation
Presented To
The Academic Faculty

By

Michael D. Kulis Jr.

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology

May 2006

Islet Neogenesis Associated Protein-Related Protein:
From Gene to Folded Protein

Thesis Committee:

Dr. Suzanne B. Shuker, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Nael McCarty
School of Biology
Georgia Institute of Technology

Dr. Donald Doyle
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Bridgette Barry
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Allen Orville
School of Chemistry and Biochemistry
Georgia Institute of Technology

ACKNOWLEDGEMENTS

God, I thank you for giving me the ability and determination to become a scientist. You could have taken me so many times, and for that I'm convinced I'm here for a real reason. Robin, I can't thank you enough for helping to turn my mind to the Most High. To my family, I love you ALL!! Mom and Grama, without you I would be nowhere. Dad, I'm glad I decided to forgive you…I love you. Chris and Kelly, what can I say, just keep doing your thing, you will find your treasures in due time. Amber, I'm just glad we met. We went through so much together, it's amazing we're still close. I will always be here for you…anything you need I will do my best to provide. Because of you I know what unconditional love is.

As for everyone at IBB: Suzy, thanks for taking a chance on me. It's been hard at times, but I'm a better person for going through graduate school. To my committee, thanks for having faith in me. You all could have kicked me out after that first candidacy exam, but you saw some potential in me. I don't think I've fully realized the potential and when I do it's going to be a sight to behold. Vicki, you were a great teacher, both in and out of the lab. Thanks for "understanding Sketchy". Bryan, thanks for all the help with the molecular biology and NMR. Best of luck in all you pursue. KJ, what can I say, I'll never forget the day you took me to the health center after I ate the banana walnut bagel, when I could barely breathe. Remember how swollen my face was? Damn, it still scares me just thinking about that. Josh and Beth, thanks for the entertainment. Dr. Gelbaum, thanks for putting up with me and all my failed NMR experiments. I think we both had faith that it would eventually work though. Charlie,

thanks for all your advice and especially for the help with my FPLC work. Foster, you're like a brother, sometimes I love ya, sometimes I wanna kill you. Thompson, it's great knowing you, even though UCLA hoops will always have the edge on Kentucky. Thanks for letting me live with you and all, I won't soon forget about you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

A, Ala	Alanine
A_{280}	Absorbance at 280 nm
ATP	Adenosine Triphosphate
□ cells	Beta cells
C, Cys	Cysteine
CD	Circular Dichroism
cDNA	Complementary Deoxyribonucleic Acid
CV	Column Volumes
D, Asp	Aspartic Acid
dI	Deionized
DNA	Deoxyribonucleic Acid
E, Glu	Glutamic Acid
E. coli	*Escherichia coli*
F, Phe	Phenylalanine
FPLC	Fast Protein Liquid Chromatography
G, Gly	Glycine
GLP-1	Glucagon-Like Peptide 1
GLUT	Glucose Transporter
GPC	Gel Permeation Chromatography
GST	Glutathione-S-Transferase
H, His	Histidine

HLA	Human Leukocyte Antigen
HPLC	High Pressure Liquid Chromatography
HSQC	Heteronuclear Single Quantum Correlation Spectroscopy
HTLV-I	Human T-Cell Lymphotropic Virus, Type 1
I, Ile	Isoleucine
IMAC	Immobilized-Metal Affinity Chromatography
IPTG	Isopropyl-□-D-Thiogalactoside
INF□	Interferon Gamma
INGAP	Islet Neogenesis Associated Protein
INGAPrP	Islet Neogenesis Associated Protein-Related Protein
K, Lys	Lysine
K	Kelvin
kD	Kilodaltons
L, Leu	Leucine
LB agar	Luria-Bertani, Miller
LB broth	Luria-Bertani, Lennox
M, Met	Methionine
MALDI	Matrix-Assisted Laser Desorption Ionization Mass Spectrometry
mg	Milligram
MHC	Major Histocompatibility Complex
mL	Milliliter
mM	Millimolar
MOE	Molecular Operating Environment

mRNA	Messenger Ribonucleic Acid
ms	Milliseconds
MWCO	Molecular Weight Cut-Off
N, Asn	Asparagine
Ngn-3	Neurogenin-3
NMR	Nuclear Magnetic Resonance Spectroscopy
NO	Nitric Oxide
NOESY	Nuclear Overhauser Effect Spectroscopy
NTA	Nitrilotriacetic Acid
$O.D._{600}$	Optical Density at 600 nm
P, Pro	Proline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PDX-1	Pancreatic Duodenal Homeobox Gene
ppm	Parts Per Million
Q, Gln	Glutamine
R, Arg	Arginine
Reg	Regenerating Protein
RNA	Ribonucleic Acid
ROESY	Rotating-Frame Nuclear Overhauser Effect Spectroscopy
RPM	Revolutions Per Minute
S, Ser	Serine
SAR	Structure Activity Relationship

SDS-PAGE	Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis
STZ	Streptozotocin
T, Thr	Threonine
τ_c	Correlation Time
T1D	Type 1 Diabetes
TFE	2.2.2-Trifluroethanol
TNF	Tumor Necrosis Factor
TOCSY	Total Correlation Spectroscopy
µL	Microliter
V, Val	Valine
W, Trp	Tryptophan

SUMMARY

Type 1 diabetes is the direct result of an autoimmune attack on the pancreatic islet cells. The islets contain β cells, which are the only type of cell capable of supplying insulin in the human body. The destruction of these cells leaves the diabetic to rely on exogenous insulin to maintain a normal blood sugar level. Without insulin injections the patient's body will be unable to utilize glucose and will break down fat for energy. The metabolism of fat leads to the production of ketone bodies, which acidify the blood, leading to ketoacidosis. Ketoacidosis is an emergency condition that can lead to coma or even death.

Insulin therapy allows the diabetic patient to deal with the symptoms of the disease, but does nothing for the underlying condition. In order to truly cure the disease, the strategy is to replenish the β cells in the diabetic. Islet neogenesis associated protein (INGAP) has been shown to regenerate islet cells and reverse experimentally-induced diabetes in animal models. The INGAP pentadecapeptide is a 15 amino acid peptide from INGAP with comparable activity to the full-length protein. This 15-mer is undergoing clinical trials for treating diabetes.

The overall goal of the project described in this work is to determine the structure of the INGAP pentadecapeptide for use in structure-based drug design of non-peptide mimics of the 15-mer. The first set of experiments in the present work directly examined the 15-mer in solution using NMR. No stable structure of the small peptide was found. The second set of experiments involved a homolog of INGAP, called INGAP-related protein, or INGAPrP. INGAPrP was recombinantly produced in *E. coli* and subsequently

purified and refolded. Refolding of INGAPrP was verified by a ^{1}H-^{15}N HSQC experiment. CD experiments supported the NMR study, indicating helical content in INGAPrP. The folded nature of the protein will allow for the total three-dimensional structure of INGAPrP to be determined. The protein structure will show the fold of the 15-mer within the full-length protein. This information will be valuable for the ultimate goal of producing structural mimics of the INGAP pentadecapeptide. Non-peptide mimics should have better oral bioavailability and longer half-lives *in vivo*.

CHAPTER 1

ISLET REGENERATION: A POTENTIAL CURE FOR TYPE 1 DIABETES

1.1 Type 1 Diabetes

There are an estimated 200 million people living with diabetes, of which some 20 million live in the United States (Zimmet et al., 2001). Insulin-dependent diabetes mellitus, or type 1 diabetes, accounts for approximately 10% of all diabetes cases worldwide, while type 2 diabetes accounts for the vast majority. Type 1 diabetes (T1D) is an autoimmune disease in which the insulin-producing cells are destroyed, leaving the patient dependent on exogenous insulin (Mathis et al., 2001). Type 2 diabetes is usually associated with obesity and is the result of a loss in the patient's sensitivity to their natural insulin (Moller, 2001). Although the prevalence of T1D is much lower than type 2, a cure is still highly sought after due to the chronic nature of the disease and the long-term complications associated with it.

T1D is usually diagnosed at a young age and is therefore also referred to as juvenile-onset diabetes. The classic symptoms include frequent hunger, urination, and thirst (The Merck Manual, 1997). Frequent hunger is the result of the body being unable to utilize glucose as a source of energy since the absence of insulin renders the cells unable to take-up glucose. These excess sugar molecules then act as an osmotic diuretic, which in turn leads to the increased thirst. T1D is commonly diagnosed by two tests. One test involves measuring a patient's fasting blood glucose levels, while glucose levels in the second test are measured after an oral glucose challenge. Diabetics will register high blood glucose levels on both tests.

The main characteristic of T1D is the elevated blood sugar levels. This hyperglycemic state results from the extremely low levels of circulating insulin in the diabetic. Insulin is one of the hormones produced in the endocrine portion of the pancreas and is secreted by the β cells in response to an increase in blood sugar levels. The glucose transporters, GLUT2, on β cells transport excess glucose into the cells, which causes a signaling cascade leading to the secretion of insulin (Saltiel and Kahn, 2001). Thus, GLUT2 acts as the glucose sensors for the human body. Without functioning β cells, the diabetic loses the ability to detect increased blood glucose levels and can not produce insulin.

Insulin is a 51 residue polypeptide consisting of two chains, A and B, connected via two disulfide bonds (Stryer, 1995). It is produced as a preprohormone. The signal peptide is cleaved in the endoplasmic reticulum and the prohormone is then packaged into secretory vesicles in the Golgi. In the Golgi, the prohormone is folded into its native state and locked into this structure by the formation of the two disulfide bonds. Protease cleavage removes the center portion of the polypeptide, known as C-peptide, which leaves the amino-terminal B-peptide bonded to the carboxy-terminal A-peptide, yielding the functional insulin molecule (Figure 1.1).

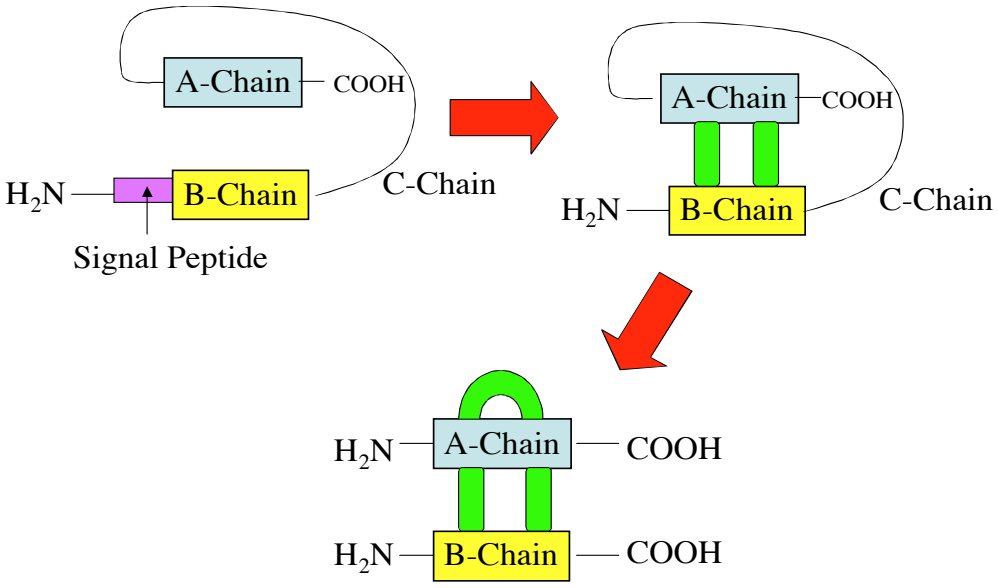

Figure 1.1 Insulin is produced as a preprohormone. Removal of the signal peptide (purple) produces proinsulin. Disulfide bond formation (green) and removal of the C-chain produce mature, biologically active insulin.

Insulin is secreted into the blood and exerts its action upon binding the insulin receptor that is found on muscle and fat tissues. The insulin receptor is a tyrosine kinase receptor and consists of two □ subunits and two □ subunits (Saltiel and Kahn, 2001). The alpha subunits are extracellular and are responsible for binding insulin, while the □ subunits are intracellular. Once insulin has been bound, the □ subunits are autophosphorylated to become active. The activated receptor then phosphorylates other intracellular proteins triggering a signaling cascade. The end result of the cascade is the fusion of glucose transporters, GLUT4, to the cellular surface. GLUT4 allows cells to bring glucose into the cell for subsequent metabolism. The lack of insulin in T1D means

that glucose can not be taken up from the blood, rendering the hyperglycemic state as well as cellular starvation.

Although the etiology of T1D is not fully understood it is clear that the disease is a direct result of reduction in β cell mass. β cells are one of the four types of cells found in the endocrine tissue of the pancreas, which contains cell clusters known as the Islets of Langerhans (Stryer, 1995). The other cell types, alpha, delta, and gamma produce glucagon, somatostatin, and pancreatic polypeptide, respectively. The islet cells constitute only 5% of the pancreas, yet are crucial for normal metabolism.

Destruction of β cells is believed to be caused by an autoimmune attack directed specifically at these cells. Researchers have shown a correlation between the human leukocyte antigen (HLA) region on chromosome six, specifically HLA-DQ and HLA-DR, and susceptibility to T1D (Lernmark, 1999; Moustakas et al., 2002). However, the presence of these genes is not sufficient to cause the disease. It is believed that an environmental factor, possibly an unidentified virus or a bacterial toxin, sets off the chain of reactions ultimately resulting in β cell depletion (Akerblom et al., 2002).

The HLA-DQ and HLA-DR genes produce antigens, which are presented by the major histocompatibility complex (MHC) on β cells. The HLA region allows the immune system to differentiate between "self" and "non-self", and in this way can produce an immune attack on cells displaying pathogenic, "non-self", antigens. However, in the case of autoimmune diseases, the immune system mistakes "self" antigens as those of a pathogen and a subsequent immune response is initiated to destroy the cells. In the T1D patient, their immune system has attacked the β cells.

The autoimmune response is associated with antibodies against antigens on the cells. There are three main cell autoantibodies, which are directed at islet cell antigen-2, glutamic acid decarboxylase (GAD65), and insulin itself, which have been correlated with T1D (Verge et al., 1998). Some researchers feel that these antigens could be an early sign of disease propensity and could be used for early treatment, or even prevention (Lernmark, 1999).

T-lymphocyte mediated cell death occurs via two pathways (Notkins, 2002). CD8+ cytotoxic lymphocytes recognize antigens bound to MHC-I on the cell surface (Figure 1.2). CD4+ helper lymphocytes recognize processed antigens bound to MHC-II molecules on the surface of antigen presenting macrophages and dendritic cells, resulting in a reaction that destroys nearby cells (Figure 1.3).

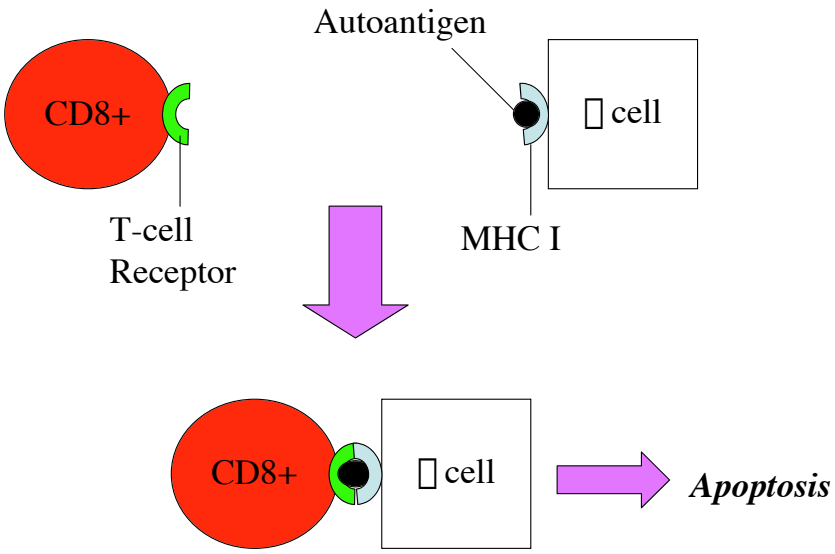

Figure 1.2 CD8+ destruction of cells occurs through the MHC I presentation of an autoantigen, which is recognized by the T-cell receptor.

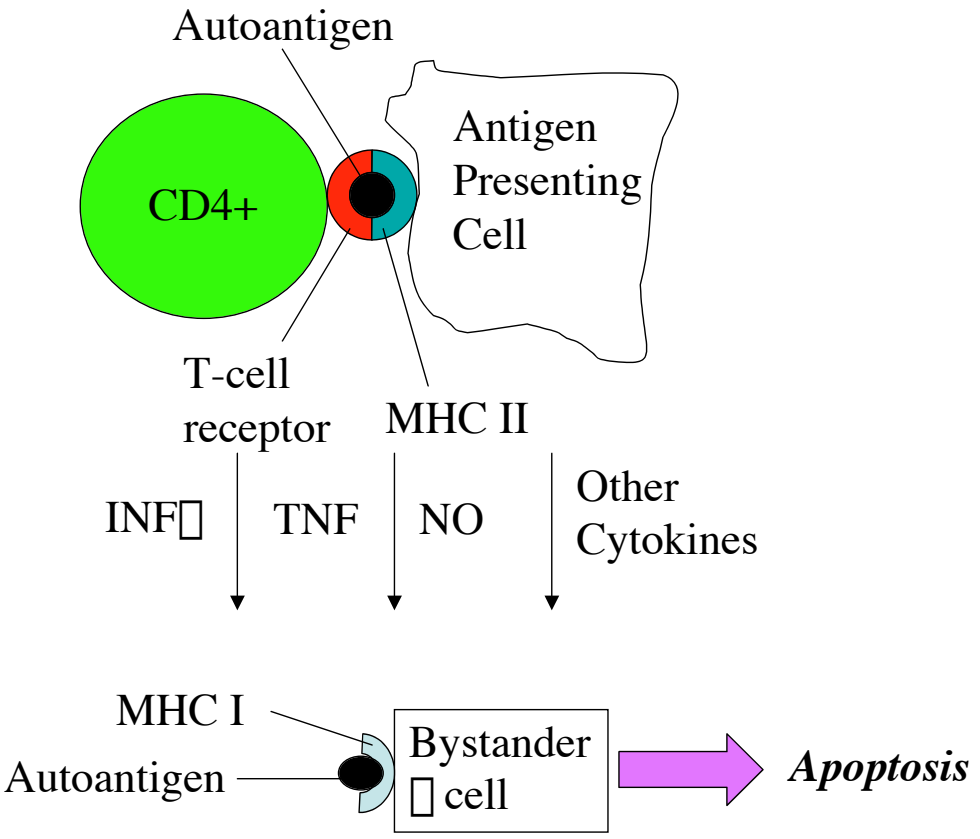

Figure 1.3 CD4+ destruction of cells occurs after the T-cell receptor recognizes an autoantigen on an antigen presenting cell. The subsequent signaling cascade leads to cell death.

Upon cell destruction, and the nearly complete loss of circulating insulin, the diabetes sufferer must turn to exogenous insulin to regulate glucose metabolism. Supplying the diabetic with insulin will allow glucose to be utilized for energy and subsequent storage of energy for later use. Without insulin injections, the diabetic's cells have no glucose to metabolize and will break down fat for energy. Utilizing fatty acids

for energy results in the accumulation of ketone bodies in the blood, rendering the blood acidic. This state, known as diabetic ketoacidosis, is an emergency condition that can result in coma, or even death (The Merck Manual, 1997).

Life-threatening diabetic ketoacidosis is not the only risk associated with T1D. T1D also causes many complications later in life due to constantly elevated blood sugar levels, which damage blood vessels and nerves (Schwartz and Bornfeldt, 2003). The build-up of complex sugar-based substances can lead to plaque build-up in blood vessels. This atherosclerosis causes poor circulation through blood vessels, which can damage the heart, brain, legs, eyes, kidneys, nerves, and skin. Some of the long-term complications associated with T1D are retinopathy, nephropathy, and neuropathy (Bailes, 2002). Retinopathy is caused by damage to the blood vessels of the retina, resulting in decreased vision, and sometimes blindness. Nephropathy is the term given to a damaged kidney. When the blood vessels in the kidney thicken, it can malfunction or even result in complete failure, rendering the patient to go on dialysis. The inadequate blood supply can also lead to neuropathy, or nerve damage. This nerve damage can render a patient's limbs weak and unable to detect simple sensations, sometimes leading to amputation.

The current protocol for a T1D sufferer is to lower the blood sugar level through daily shots of insulin. The patient must monitor his or her blood sugar levels before and after meals and administer the insulin when appropriate (The Merck Manual, 1997). This is difficult for people to do and still live a normal life. It is especially hard on children afflicted with T1D, whose parents must constantly watch over them. The shortcomings of insulin therapy have brought researchers to explore new T1D therapeutics.

1.2 Therapeutic Strategies for Type 1 Diabetes

The unattractiveness of daily insulin injections and skin pricks to check blood sugar levels, as well as the imperfect results of blood sugar regulation via insulin injections, brings one to the conclusion that it would behoove the medical community to discover alternative approaches to treat T1D. There are several strategies currently being studied, including islet transplantation, gene therapy, insulin mimics, and islet cell regeneration (Yamaoka, 2003).

Islet transplants are now greater than 80% effective thanks to the advent of the Edmonton Protocol (Shapiro et al., 2000). The Edmonton Protocol uses pancreatic material from two cadaveric donors to transplant enough functioning β cells into the portal vein of the liver (Shapiro et al., 2000). This location is optimal for transplantation because of its seclusion from autoantibodies directed against β cell antigens. Unfortunately, the procedure relies on a life-long regimen of immunosuppressive drugs. These drugs leave the body with a compromised immune system that often leads to opportunistic infections. However, the main problem with this therapy is the severely limited number of donor pancreases. Only ~3000 organs become available in the United States each year, and the technique requires islets from two donors for each transplant. This is obviously not nearly enough for the two million type 1 diabetes sufferers.

Gene therapy strategies for T1D involve integration of the insulin gene under control of a specific promoter, which will be activated upon increased glucose concentration (Yoon, 2002; Fenjves, 2000). Gene therapy has the advantage over other approaches in that it will not fall victim to a repeat autoimmune attack. This method has shown promise in animal models, however, there are several barriers. Regulation of the

transgenic insulin production has been problematic, meaning that either too little or too much insulin is produced (Kolodka, 1995; Tuch, 1998). Not enough insulin means that the diabetic will experience high blood sugar levels, and too much insulin can lead to a potentially lethal state of hypoglycemia. As with all gene therapy approaches, another concern is how to deliver the gene to the desired location within the human genome. These problems will have to be overcome before the gene therapy approach is a viable T1D treatment option.

A specific example of using gene therapy as a strategy to reverse T1D was accomplished by researchers at Emory University (Thule and Liu, 2000). Their system relied on hepatic production of insulin following gene transfer to hepatocytes. Insulin expression was controlled by glucose responsive elements from the pyruvate kinase gene, thus acting as the glucose sensor. This method was shown to be effective in maintaining normal blood sugar levels following glucose challenges in STZ diabetic rats. This study offers hope for the gene therapy paradigm of T1D treatment.

Insulin mimics are another avenue being explored to treat T1D (Qureshi et al., 2000; Ding et al., 2002). An insulin mimic is a small molecule that will bind to and activate the insulin receptor, thereby translocating GLUT4 to the cellular surface to provide up-take of glucose into the cell. Insulin mimics have the advantage over insulin in that they would be orally available as opposed to insulin, which must be injected. However, patients would still have to monitor blood sugar levels before and after meals and take an insulin mimic accordingly.

Islet regeneration therapy holds a distinct advantage over gene therapy and insulin mimics in that islet regeneration would treat the underlying cause of T1D, whereas the

other two alternatives only treat the symptoms (Yamaoka, 2002). Islet regeneration can be done *in vitro* with subsequent transplantation of the islets, or it can be done *in vivo*. Either way the patient is provided with a functioning mass of cells to regulate blood sugar levels. The main drawback of islet regeneration is that subsequent destruction of the newly produced cells will likely occur via the same mechanism that originally caused T1D. Therefore, researchers must think in terms of a two-prong approach in which new islets are introduced in a patient and the immune attack directed at the cells is halted.

1.3 Regenerative Medicine

Regenerative medicine can be thought of as a way to repair or replace failing human tissue or organs. The general strategy involves harvesting pluripotent cells, growing the cells in the lab, and stimulating the cells to form specific tissues (Heber-Katz, 2004). These tissues are then examined and shown to be effective before placement into a diseased body.

Alternatively, pluripotent adult stem cells can be differentiated within the human body (Rosenberg 1995). This however would theoretically only provide temporary relief of the disease because whatever caused the disease in the first place would happen again. In this sense, *in vivo* regenerative medicine is at a disadvantage because with *in vitro* regeneration researchers can modify the cells in such a way to prevent a reiteration of disease progression. *In vivo* regeneration is also difficult because the factors responsible for differentiating pluripotent cells are often peptides or proteins that are prone to swift degradation in the body.

Although there are some hurdles yet to be cleared, regenerative medicine gives hope to patients with a myriad of diseases. The diseases hoped to be curable via regenerative medicine include Parkinson's disease, Alzheimer's disease, arthritis, osteoporosis, emphysema, liver cirrhosis, heart disease, cystic fibrosis, and diabetes (Herber-Katz, 2004).

T1D is an attractive target for regenerative medicine because it involves replacement of only one type of cell, namely the pancreatic β cell. Replenishment of insulin-producing β cells within a diabetic patient coupled with a therapy to prevent destruction of the new β cells could reverse the disease state (Sumi and Tamura, 2000). Thus, regenerative medicine offers real hope to T1D sufferers.

1.4 Islet Regeneration

Islets are known to have the ability to regenerate (Rosenberg, 1995; Wilkin, 1998). Research conducted from 1902 to 1914 showed that islet and acinar regeneration occurred after pancreatic duct obstruction (Shaw et al., 1926). In 1924, these same cells were observed to regenerate after near total pancreatectomy (Shaw et al., 1926). In 1947, Hughes found that pancreatic injury, resulting from administration of a β cell toxin, alloxan, in rats caused an initial loss of islets but then noticed an increase in islet number in a subsequent regenerative phase (Hughes, 1947). This same phenomenon was observed in the 1980s when streptozotocin, a β cell toxin, was administered to neonatal rats (Cantenys et al., 1981).

There are two hypotheses for the mechanism of regeneration *in vivo* (Bouwens, 1998). One idea is that new islets are spawned from residual stem cells. A second

hypothesis is that new islets are produced from proliferation of existing cells (Dor et al., 2004). Both of these mechanisms may be viable pathways to produce insulin-expressing cells. Although the debate continues as to the mechanism of regeneration, one thing is certain: finding the factors responsible for producing new islets could lead to a cure for T1D.

The first protein isolated from regenerating pancreata was the Reg, or regenerating, protein (Terazono et al., 1988). Researchers produced an animal model of diabetes in which a diabetic state was induced by 90% pancreatectomy. The researchers then began treatment with the poly(ADP-ribose) (PARP) synthetase inhibitor nicotinamide. Overproduction of PARP can lead to tissue necrosis, so the hypothesis was that nicotinamide could stop cell death, leading to regeneration of the pancreatic tissue. Exocrine and endocrine cell regeneration was noticed with the presence of the Reg protein. This regenerating factor was expressed only in regenerating pancreata and not in normal animals. The protein was shown to improve the diabetic state in an animal model as evidenced by a significant decrease in blood glucose levels and an increase in cell mass (Watanabe et al., 1994).

Reg has also been seen in other models of islet regeneration. A method to reverse cell damage was investigated by administration of an antioxidant, probucol, to STZ diabetic hamsters (Takatori, 2003). The hypothesis was that oxidative damage to islets was key in the pathogenesis of diabetes. The researchers found that upon administration of probucol, Reg and islet neogenesis associated protein (INGAP) gene expression were upregulated. The study concluded that there was partial restoration of cell function, likely due to the increased presence of Reg and INGAP.

Since the Reg protein was identified, many other factors have been discovered which cause islet generation both *in vivo* and *in vitro*. Some of these include glucagon-like peptide-1 (GLP-1) (Hui et al., 2001; Buteau et al., 2003), exendin-4 (U.S. Department of Health and Human Services, 2000), dipeptidyl protease IV inhibitors (Probiodrug, 2002), islet neogenesis associated protein (Rafaeloff et al., 1997), rat pancreatic extract (Soo Kim 2003), epidermal growth factor (EGF) (Brand et al., 2002), gastrin (Brand et al., 2002), transforming growth factor-□ (Sandgren et al., 1990), betacellulin (Watada et al., 1996), activin (Mashima et al., 1999), insulin-like growth factor-I (IGF-I) (George et al., 2002) and interferon-gamma (Gu 1993). Several review articles have been written that cover these regeneration factors (Yamaoka and Itakura 1999; Nielsen et al., 1999; Sumi and Tamura, 2000; Soria, 2001; Risbud and Bhonde, 2002; Yamaoka, 2002). It should be pointed out that some of these factors may not have to be administered to the patient but could be produced *in vivo* through use of gene therapy to deliver the gene encoding the protein factor of interest, leading to subsequent expression and islet regeneration (Yamaoka 2001),

Other factors of importance include transcription factors, such as the pancreatic duodenal homeobox gene, PDX-1, also referred to as IDX-1 (Sumi and Tamura, 2000; Gagliardino et al., 2003). This gene is a requisite for islet generation from stem cells. It is believed to be active in the normal pancreatic development during the fetal stage. Another transcription factor is neurogenin-3 (Ngn-3), which is also key in islet differentiation (Vetere et al., 2003). Researchers hope to exploit this knowledge by identifying cells with these transcription factors as candidates for islet generation. One

can imagine developing cell lines with high levels of PDX-1 or Ngn-3 for mass production of islet cells, or targeting PDX-1/Ngn-3 expressing cells *in vivo*.

As already mentioned, the paradigm of islet regeneration consists of two steps. The first step is to generate insulin-expressing cells *in vivo*, or to generate the cells in the lab and then transplant them into the patient (i.e. Edmonton Protocol). This step has been explored above. The second step is to prevent these newly supplied cells from being attacked by the immune system. Researchers have come up with several ways to combat the autoimmune response (Cotterell and Kenyon, 2002).

Encapsulation of islet cells produced in the lab prior to transplantation could prevent immune destruction (Gray, 1997; Maria-Engler et al., 2001). Cells can be encapsulated by various polymers, which strategically allow influx of nutrients and outflow of insulin, but prevent the up-take of antibodies directed at the islets. The polymeric thickness can be controlled so that encapsulated cells will have a wide range of diffusion properties, potentially allowing for fine tuning of insulin release.

Another way to prevent immune attack of islet cells is to modify their cell surfaces so that they will not display the aforementioned cell autoantigens (Yamaoka, 2001). Manipulating the cells' properties prior to transplantation should be a straightforward and powerful tool to reduce the chances of a recurrent immune attack.

A third method researchers hope to develop that will spare transplanted islets is by targeting the immune system itself. This is being investigated via monoclonal antibody therapy (Keymeulen et al., 2005). Monoclonal antibodies will bind to the autoantibodies directed at the diabetic's islet cells, thus destroying the autoantibodies' function.

With the advent of so many methods to produce islets *in vivo* and *in vitro*, and new methods of protection from autoimmune response, it seems that it is only a matter of time before T1D can be cured by islet regeneration therapy. Figure 1.4 illustrates the necessary steps to treat T1D via *in vitro* islet regeneration.

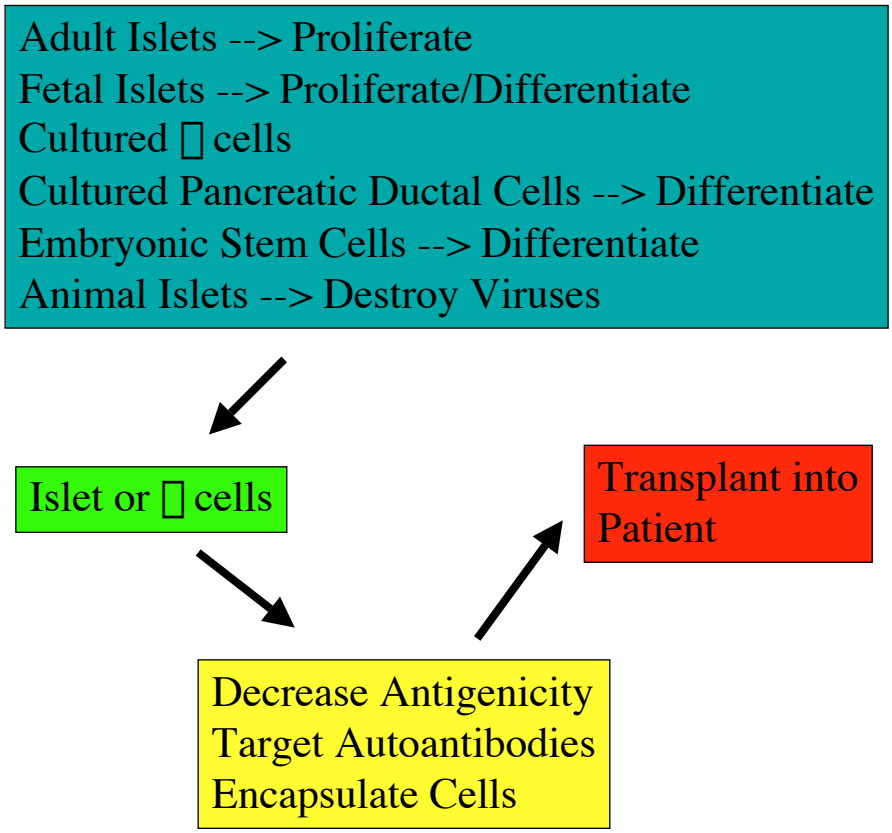

Figure 1.4 Summary of T1D treatment by islet regeneration.

CHAPTER 2

INGAP AND INGAPrP

2.1 INGAP: Discovery and Activity

The search for INGAP started in 1983, when researchers found that wrapping the head of a hamster pancreas caused islet cell growth (Rosenberg et al., 1995). The experiment was intended to induce pancreatitis, but instead of inflammation the researchers noticed new islet formation. The researchers termed this the "partial duct obstruction" model of islet regeneration (Rosenberg, 1998).

Rosenberg's method of partial duct obstruction allowed for subsequent islet formation to be studied without pancreatitis, autoimmune destruction, or tissue atrophy. The islet cell neogenesis did not appear to be mitotic division of existing cells, but rather reiteration of normal ontogeny from ductular cell precursors. Immunocytochemical techniques confirmed mature islets were produced that expressed insulin, glucagon, and somatostatin (Rosenberg, 1998).

The ultimate test for Rosenberg's model of islet regeneration was to evaluate the cellophane wrapping technique to see if it could reverse a diabetic state. Using hamsters treated with streptozotocin (STZ), and therefore diabetic, the test was passed (Rosenberg, 1998). New islets formed and the diabetic state was reversed in greater than 50% of the animals. These animals now had normal serum glucose and insulin levels. The next step was to determine what was causing this reversal.

Using standard fractionation and protein chemistry techniques, an extract, termed Ilotropin, was identified in the regenerating pancreata that was not in normal pancreata

(Pittenger et al., 1991). Ilotropin was found to be protease sensitive, acid stable, heat stable, alcohol precipitable, and had a molecular weight range of 29-44 kD. It was apparent this compound was a protein, however it appeared to be a mixture of several proteins.

The researchers, led by Dr. Vinik and Dr. Rosenberg, turned to another strategy to firmly establish which factor was causing this unexpected reaction. They used a technique called mRNA differential display (Liang and Pardee, 1992). The technique uses poly dT affinity chromatography to isolate mRNA from various cells or organs and then allows the user to compare which mRNAs are present in the subsets of cells. For example, the technique can be used to look at which mRNAs are present in a lung cancer cell but not present in a normal lung cell. This gives researchers a method to elucidate causes of a disease and to determine which proteins may be valid targets for therapeutic intervention.

In the case of INGAP, researchers used mRNA differential display and identified ten mRNA transcripts that were uniquely or over-expressed in the regenerating pancreata but absent, or expressed at very low levels, in normal hamster pancreases (Rafaeloff et al., 1996). One clone in particular, RD 19-2, was found to be of extreme interest. RD 19-2 had some homology to the Reg gene, which instantly made it a candidate to be the cause of islet regeneration. It was also found to be maximally expressed at 1-2 days following cellophane wrapping of the pancreas without expression at any other time, and was not observed in normal pancreata. The researchers reasoned that the gene product may be upregulated following the apparent pancreatic injury, which would lead to a

regeneration of islet mass to compensate for the tissue trauma caused by the cellophane wrapping.

RD 19-2 was later termed the INGAP gene (Rafaeloff et al., 1997). The name is an acronym for islet neogenesis associated protein. The mRNA transcript is 766 bp and the encoded INGAP protein is 175 amino acids long. Tissue-specific expression was determined via Northern blotting. It was demonstrated that INGAP RNA was present in pancreas and duodenum, but not observed in spleen, stomach, lung, liver, heart, or muscle. A Western blot using antibodies raised against INGAP demonstrated that INGAP was a component of the earlier protein mixture, Ilotropin, identified by the researchers. Thus, the mRNA differential display proved successful.

The next test was to see if this protein could reverse experimental diabetes in an animal model. The researchers cloned the gene into an expression vector for recombinant production in *E. coli*. Figure 2.1 shows the amino acid sequence of INGAP. It was found that in order to express the protein, the researchers needed to exclude the signal peptide, which is the first 26 amino acids. Upon expression, the protein was purified by use of the his-tag at the N-terminus (Vinik et al., 1998). The protein was then dialyzed into a Tris buffer at pH 8.5 for *in vitro* and animal testing.

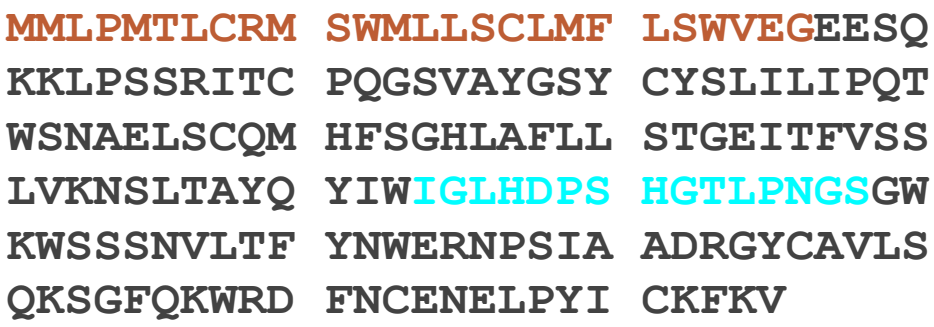

Figure 2.1 INGAP sequence. The first 26 residues constitute the signal peptide and are shown in brown. The INGAP pentadecapeptide is shown in blue.

In vitro testing was done with the ARIP cell line, which is isolated from rats and is a ductal epithelial cell line. Dr. Vinik's team incubated ARIP cells with INGAP and measured the [^{3}H]-thymidine uptake from the media (Rafaeloff et al., 1997). [^{3}H]-thymidine was indeed taken-up from the media, indicating that the cells were dividing. This was an encouraging finding since proliferation of ductal cells is a prerequisite for islet neogenesis.

STZ-induced diabetic mice and hamsters were chosen as animal models. INGAP was found to be active and could reverse STZ-induced diabetes in greater than 50% of the test animals (Gold et al., 1998). Quite an accomplishment, however, the researchers knew that taking a relatively large protein like INGAP into clinical trials may be difficult because it would be rapidly destroyed by proteolytic activity *in vivo*. They strived for a better compound to take into the clinic. The INGAP sequence was searched, and the researchers found a region that was promising for biological activity.

The stretch of amino acids from 104 to 118 (Figure 2.1) within INGAP had potential biological activity for two reasons, according to the researchers (Rafaeloff et al.,

1997). This region of INGAP differs from Reg because of a unique insertion of five amino acids, SHGTL. It was apparently believed that this may make INGAP somehow more active than Reg. The second reason is that the stretch of amino acids precedes a potential N-glycosylation site at position 126, thus indicating potential biological activity.

This pentadecapeptide, IGLHDPSHGTLPNGS, was synthesized and shown to have activity in reversal of diabetes in animals (Rosenberg et al., 2000). This smaller version of INGAP should be easier to administer, have a longer half-life *in vivo*, and cause fewer side effects.

The peptide nature of the 15-mer means it would still have to be injected into patients and would be broken down rather quickly. It would therefore be more desirable to have a non-peptide compound with the same activity as the 15-mer. We thought about this and wanted to use our lab's expertise in protein NMR to solve the structure of the 15-mer and use that information to develop better therapeutics. As will be shown in Chapter 3, the 15-mer has no stable structure in solution under the conditions we tested. This led us to believe that the 15-mer needed the surrounding protein to give it structure. Therefore, we wanted to solve the INGAP structure and use that to come up with synthetic targets for peptide mimetic drugs of the 15-mer. We could not acquire INGAP, so we used a homolog, INGAP-related protein (INGAPrP).

2.2 INGAPrP

Researchers in Japan wanted to identify the mouse version of INGAP. What they found was islet neogenesis associated protein-related protein (INGAPrP). INGAPrP mRNA was isolated from mouse and the predicted protein sequence was found to be 72%

identical to INGAP (Figure 2.2) (Sasahara et al., 2000). It was found by using the primers for the INGAP gene to produce a PCR product from mouse duodenum cDNA. The PCR product was sequenced and a mouse cDNA library was searched to determine if the gene was already in a gene bank. The gene was found in a mouse cDNA library and had the accession number AA822059.

```
INGAP     MMLPMTLCRMSWMLLSCLMFLSWVEGEESQKKLPSSRITCPQGSVAYGSY
INGAPrP   MVSHKTLHSMSWMLLCCLMSLSWVQGEQSQKKLSSPRISCPQEAQAYGSY

INGAP     CYSLILIPQTWSNAELSCQMHFSGHLAFLLSTGEITFVSSLVKNSLTAYQ
INGAPrP   CYLLILEPQTWANAEIHCQKHFSGHLAFLLTYGEIIFVSSLVKNSLTTFP

INGAP     YIWIGLHDPSHGTLPNGSGWKWSSSNVLTFYNWERNPSIAADRGYCAVLS
INGAPrP   YIWIGLHDLSLGSLPNENGWKWSSSDPLTFYNWEIPPSMSAHHGYCAALS
             ***** * * ***

INGAP     QKSGFQKWRDFNCENELPYICKFK
INGAPrP   QASGYQKWRDYYCDPTFPYVCKFK
```

Figure 2.2 Sequence alignment of INGAP and INGAPrP. The proteins are 72% identical. Identical residues are shown in pink. Identical residues within the INGAP pentadecapeptide region are underscored with blue stars.

INGAPrP was detected in various tissues throughout the mouse (Sasahara et al., 2000). Northern blot analysis confirmed INGAPrP's presence in the pancreas and duodenum, just like INGAP. However, INGAPrP was also detected in the stomach and skeletal muscle. This was an unexpected result for the researchers because INGAP was believed to be exclusively produced in regenerating pancreatic tissue. Thus, the researchers concluded their paper by essentially saying the gene was not likely to be

useful in islet regeneration since it was detectable in normal pancreata and various other mouse tissues (Sasahara et al., 2000). However, it was later demonstrated that INGAP is present in normal pancreases as well (Flores et al., 2003). This leads us to believe that INGAPrP may have regenerative capabilities after all.

Regardless of the activity of INGAPrP, structural studies of the protein would be useful for gaining information about the INGAP pentadecapeptide's structure. INGAPrP has a putative pentadecapeptide region very similar to INGAP's pentadecapeptide (Figure 2.3). Knowledge of this homologous region's structure would be useful for making mimics of the INGAP 15-mer. It would also be interesting to know the structure of the protein as a whole, which may lead to further insight into islet regenerative therapeutics.

INGAP: ….^{104}**IGLHD**P**S**H**G**T**LPN**GS^{118}….
INGAPrP: ….^{104}**IGLHD**L**S**L**G**S**LPN**EN^{118}….

Figure 2.3 Sequence identity between INGAP's pentadecapeptide and INGAPrP's putative pentadecapeptide. Identical residues are bold and underlined.

2.3 The Reg Gene Family

INGAP belongs to the Reg gene family, which consists of members from several species, including human, rat, mouse, hamster, and cow (Okamoto, 1999). They have been classified as either Reg I, Reg II, or Reg III proteins based on sequence homology. They are all believed to be involved in some facet of islet regeneration.

Reg I proteins include: human Reg Iα, also known as pancreatic stone protein (PSP), lithostathine, and pancreatic thread protein (PTP); human Reg Iβ; rat Reg I; and mouse Reg I. The only Reg II protein is mouse Reg II, which has a unique sequence of

seven amino acids at its N-terminus. The Reg III proteins are: HIP, which is a gene expressed in hepatocellular carcinoma, intestine, and pancreas; pancreatitis-associated protein (PTP); rat Reg III, PAP I, PAP II, and PAP III; mouse Reg III□, □, □, and □; hamster INGAP; and cow PTP. INGAPrP was discovered after these designations, but would belong in the mouse Reg III family.

A receptor for Reg I was identified from a rat islet cDNA library (Kobayashi et al., 2000). The gene for the receptor encodes a 919 amino acid protein. The sequence suggests that the protein is a type II transmembrane protein with a long extracellular region, a transmembrane segment, and a short intracellular region at its amino-terminus. Upon expression of the cDNA, it was found that rat Reg I bound to the receptor with affinity in the low nanomolar range. The Reg receptor has 97% identity to human exostoses (EXT)-like gene 3 (EXTL3) and human EXT-related gene 1 (EXTR1), signifying that the rat Reg receptor is a homolog of these human genes.

No reports have appeared in the literature regarding any studies of INGAP binding or activating these receptors, and it is currently unknown which receptor INGAP binds. INGAP is believed to bind and activate a receptor since it is found in the acinar cells and acts in a paracrine manner on the pancreatic ductal cells (Rafaeloff et al., 1997). Knowing which receptor INGAP binds could provide a direct therapeutic target for islet regeneration, and may allow researchers to manipulate the structure of potential drugs, such as the INGAP pentadecapeptide, to increase their potency.

The only protein from the Reg family with a structure in the PDB is Lithostathine. Its structure was determined to 1.3 angstroms via x-ray crystallography. INGAP and INGAPrP have 42% identity to Lithostathine, which can theoretically be relevant for

homology modeling. However, it is always better to have the structure determined empirically. Nonetheless, we created a homology model that will briefly be discussed in Chapter 5.

2.4 Significance of Present Work

The present work provides a protocol for recombinant expression of INGAPrP in *E. coli*. Methods of purification, refolding, and ^{15}N-labeling of INGAPrP are also presented. ^{1}H-^{15}N HSQC and CD experiments indicate that our protocol provides a folded protein, which will allow for structure determination. Upon solving the three-dimensional structure of INGAPrP it is believed that crucial knowledge regarding the pentadecapeptide region will be gathered. Structural insight can provide useful target information for drug development. In summary, this work is the first reported expression and isolation of INGAPrP.

CHAPTER 3

NMR STUDIES ON THE INGAP PENTADECAPEPTIDE

3.1 Purpose of the NMR Studies

Two-dimensional NMR was used to study the structure of the INGAP pentadecapeptide. Determining the fold of the 15-mer will provide insight into the active conformation of the peptide. With the knowledge of the pentadecapeptide's conformation, it is reasonable to believe that peptide mimetics could be designed. This would lead to non-peptide therapeutics that should have longer half-lives *in vivo*, better oral bioavailability profiles, and may be more active.

3.2 Sample Preparation and Data Collection

The peptide was synthesized and purified by SynPep. It was prepared with acylation of the N-terminus and amidation at the C-terminus, yielding Ac-IGLHDPSHGTLPNGS-NH_2. The molecular weight was verified to be 1543.0 daltons by MALDI. Purity was determined via HPLC. The end result was a white powder, which was received in our laboratory.

The 15-mer was determined to be soluble in a phosphate buffered solution. A phosphate solution is suitable for NMR studies because there are no ^1H's from the buffer molecules to interfere with the spectrum. However, since the peptide is in an aqueous solution, a water suppression pulse sequence was needed to minimize the water signal, thus maximizing the peptide signal. WATERGATE was used to accomplish the water suppression.

All data were collected on a Bruker AVANCE DRX 500 NMR spectrometer. The experiments to perform for determining the structure of a small peptide are: one-dimensional spectra, TOCSY, and ROESY (or NOESY). The one-dimensional spectra are useful to optimize the parameters of the sample, such as concentration, pH, and temperature. The first two-dimensional experiment to collect is the TOCSY, or total correlation spectroscopy. This experiment works by performing a "spin-lock" in the pulse sequence that allows resonance to travel from the amide proton of a residue all the way down that residue's side-chain. TOCSY is thus a through-bond experiment and allows for assignment of which proton resonance frequencies belong to which proton in the peptide. ROESY is a through-space experiment and is critical for structure determination because it determines which protons are close to each other in three-dimensional space. ROESY was chosen over NOESY because of the molecular weight of the INGAP pentadecapeptide, which affects its correlation time, or □. ROESY's pulse sequence has a "mixing time" in which energy is transferred between protons within five angstroms of each other in three-dimensional space.

3.3 Results

One-dimensional spectra were collected on a 5 mM solution in 10 mM phosphate buffer. The pH of the samples were varied from 5.5 to 7.0. The spectra showed good resolution of the proton signals, however some had less overlap than others. From visual inspection it was determined that the sample in pH 6.0 buffer gave the best signal resolution. The temperature was also varied and optimized. It was determined that low temperatures gave better resolution. 274 K was chosen as the optimal temperature.

Figure 3.1 shows the amide region of the one-dimensional spectrum collected at pH 6.0 with a temperature of 274 K. These are the conditions that the two-dimensional experiments would be carried out at.

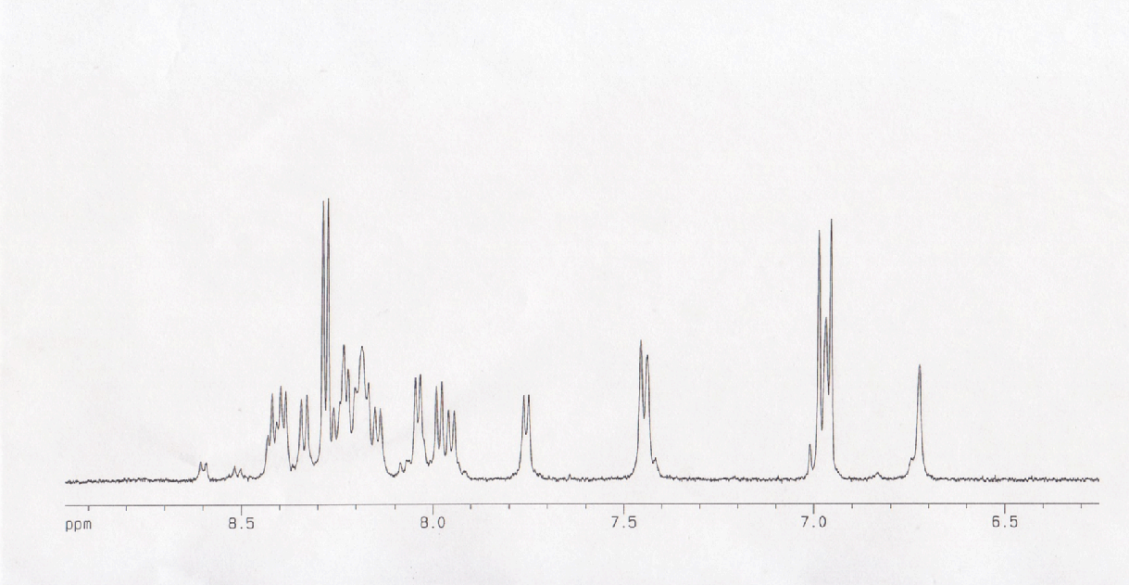

Figure 3.1 Amide region of 5 mM INGAP pentadecapeptide in 10 mM phosphate buffer at pH 6.0 and 274 K.

TOCSY data was collected on the same sample. A mixing time of 80 ms was shown to give the best spectrum. The spectrum used to assign the resonances is shown in Figure 3.2. The figure shows the amide region on the x-axis and the other protons (alpha, beta, gamma, etc) on the y-axis. The spectra on the left and top of the figure are from the one-dimensional experiment for those regions. Vertical lines illustrate where the energy transfer has occurred in the two-dimensional experiment, each line correlating to a single amino acid. Table 3.1 shows the chemical shifts of the various protons for most residues. Although not a complete assignment, this was enough information to proceed to the

ROESY experiment. It should also be pointed out that the proline residues do not have any amide protons and therefore were not assigned.

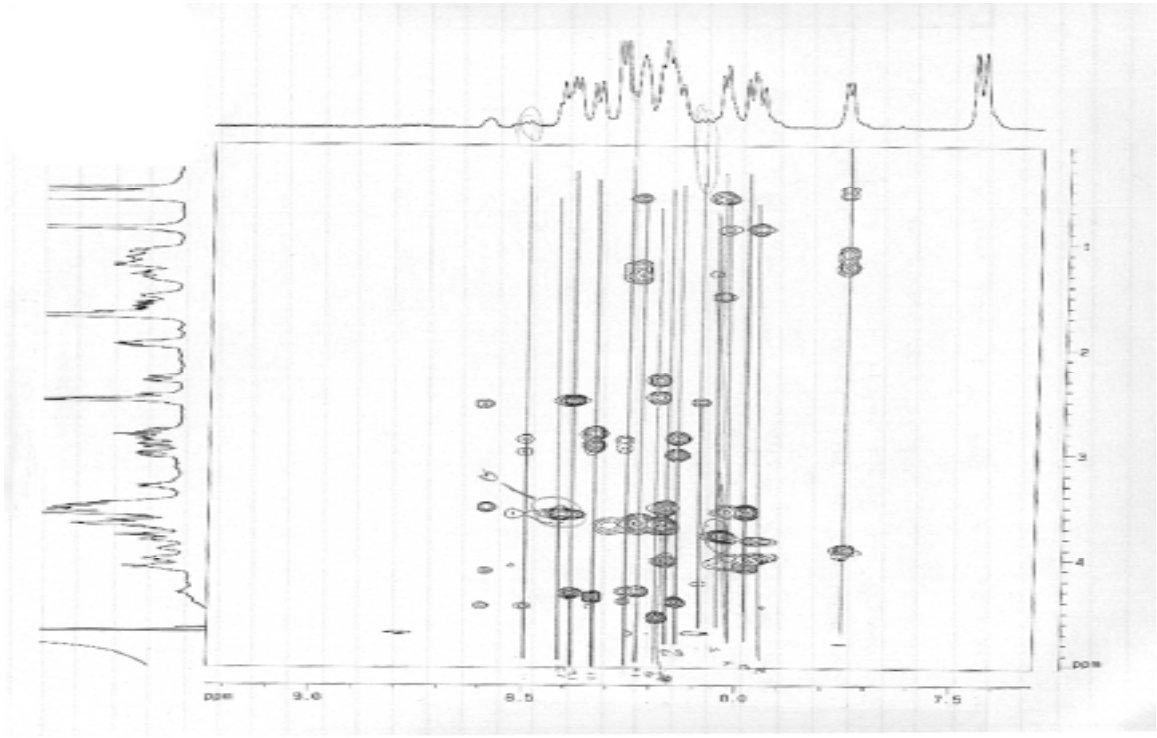

Figure 3.2 TOCSY spectrum of INGAP pentadecapeptide. The vertical lines indicate energy transfer along an amino acid's side-chain.

Table 3.1 Assignment of proton frequencies from TOCSY experiment.

N-H	□-H	□-H and others	Amino Acid
7.743	3.898	1.225, 1.091, 0.566, 0.480	Leu
7.963	3.996	3.837, 0.871	Thr
7.975	4.069	3.556	Ser
8.036	3.776	1.506, 0.872, 0.580	Ile
8.183	3.678		Gly
8.146	4.411	3.007, 2.861	His, Asp, or Asn?
8.183	4.020	3.520	Ser
8.244	3.654		Gly
8.183	4.557	2.470, 2.311	His, Asp, or Asn?
8.244	4.313	1.298, 0.578	Leu
8.329	4.362	2.922, 2.812	His, Asp, or Asn?
8.390	4.338	2.494	His, Asp, or Asn?
8.415	3.569		Gly

The ROESY spectrum is shown in Figure 3.3. This shows the full two-dimensional spectrum of data collected with a mixing time of 250 ms, which was determined to be the optimal mixing time. The results were used for some sequential assignment of amide and alpha protons (green circle, Figure 3.3) to trace the backbone of the peptide, as well as verify the assignments from the TOCSY experiment. However, the small number of inter-residue crosspeaks does not bode well for structure determination.

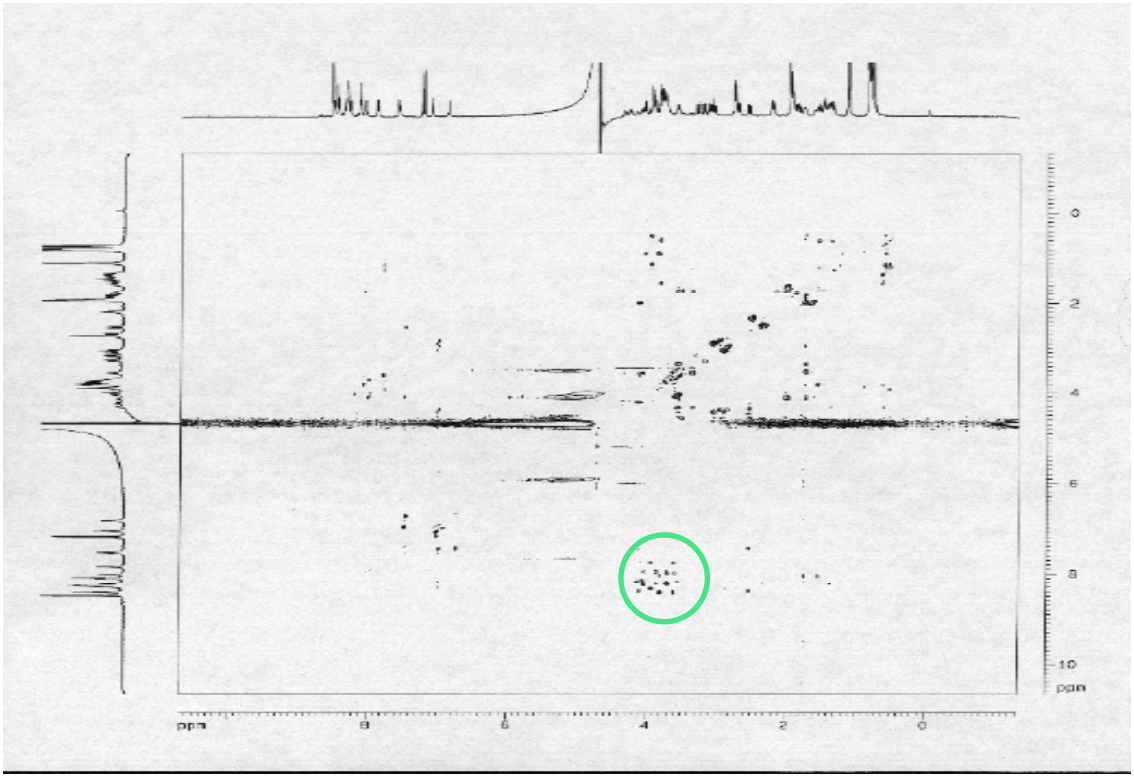

Figure 3.3 ROESY spectrum of INGAP pentadecapeptide. Peaks within the green circle were used for sequential assignment.

3.4 Discussion

The one-dimensional experiments proved useful in determining the conditions to use for the two-dimensional experiments. The TOCSY data were encouraging and easy to interpret. Assignment of proton frequencies was achieved in a timely manner.

The ROESY experiment was a disappointment. Although there were peaks observed in the amide-alpha region, there were very few inter-residue crosspeaks. The amide-alpha region allows for sequential assignment, which is useful in assignment of proton resonances. The other regions are the more informative parts of the spectrum because they show which protons are interacting in space. These interactions are absolutely crucial for structure determination because they allow for determination of constraints when the data is entered into a computer modeling program. Without a significant number of these inter-residue interactions there is no way to determine a structure. Various mixing times were tried, as well as altering the pH, concentration of the peptide, and temperatures without success. NOESY data was collected as well, but did not provide any more information than the ROESY.

The ROESY experiments essentially showed that the peptide had no stable conformation in solution. In hindsight, this was not very surprising. The peptide is a very small portion of the total INGAP protein, and since it is only a 15-mer it is unlikely to show a specific fold. The peptide also contains two glycines and two prolines. These residues are at two extremes in terms of peptide structure. Glycine residues allow free rotation around the residue because of the absence of any side-chain carbons. Prolines lock a peptide into a “kinked” conformation because of the side-chain folding back on

itself at the amide nitrogen. This severely restricts the conformations the peptide can adopt.

In light of these results it is concluded that the INGAP protein likely holds the pentadecapeptide into a stable, active conformation. Homology modeling also points to this conclusion (see Chapter 5). It can also be reasoned that the peptide is active *in vivo* even though it has no defined fold because it may bind to a receptor and subsequently adopt a particular conformation which activates the receptor. Activation of the receptor would then trigger a signaling cascade giving the 15-mer its proliferative and differentiating properties *in vivo*. This hypothesis is also supported by the fact that the full-length protein is 55-times more active than the 15-mer alone, meaning that the protein may lock the active portion into an active conformation.

CHAPTER 4

CLONING, EXPRESSION, AND PURIFICATION OF INGAPrP

4.1 Purpose of Recombinant Production of INGAPrP

INGAPrP production and isolation was desired to perform structural studies on the protein. The specific area of interest was the putative pentadecapeptide region, amino acids 104-118. This area is very similar to the INGAP 15-mer which is known to have proliferative and differentiative function on pancreatic ductal cells. Therefore, this portion of INGAPrP is of particular interest from a structural standpoint. Since the INGAP pentadecapeptide was found to have no stable structure via two-dimensional NMR experiments, we decided to study the entire INGAP-related protein (INGAPrP).

4.2 Construction of pINGAPrP Expression Vector

Standard molecular biology techniques were employed to produce the INGAPrP expression vector. The general strategy consisted of three parts. The first phase was to produce the INGAPrP gene via PCR. The second step was to digest the PCR product and the expression vector with the same restriction enzymes. The final step in the scheme was to ligate the gene and the plasmid to give the expression construct. The resulting construct was named pINGAPrP.

The INGAPrP gene was purchased from ResGen. It was from a mouse cDNA library and has the accession number AA822059 (Sasahara et al., 2000). The signal peptide of the protein was omitted from the PCR scheme because the INGAP signal peptide was reported to be toxic to *E. coli* (Vinik et al., 1998). We used the website

http://www.stepc.gr/cgi-synaptic/sigfind to predict the signal peptide from the INGAPrP protein. The first 26 residues were classified as belonging to the signal sequence. This is the same section of the full-length INGAP protein that was determined to be the signal peptide (Vinik et al., 1998).

For the expression system, we chose Novagen's pET expression system (pET System Manual, 1997). This system uses a T7 RNA polymerase, which is inducible by IPTG. Once produced, the T7 RNA polymerase binds to the T7 promoter on the pET expression construct, pINGAPrP, to produce large quantities of the pINGAPrP mRNA. The mRNA is then translated into protein by the *E. coli* ribosome.

The pET vector chosen was pET 28b. This vector is selectable by kanamycin, which makes it easy to identify and use the resulting pINGAPrP construct. The pET 28b vector also has a hexahistidine-tag gene at the 5'-end of the multiple cloning site. This allowed for insertion of the INGAPrP gene into the multiple cloning site in such a way that the resulting protein would have the 6X-His tag at its N-terminus. The fusion tag should allow for easier purification.

The PCR was performed with Vent DNA Polymerase (New England Biolabs) and yielded a fragment with a 5' NdeI site and a 3' BamHI site. Again, this omitted the signal peptide. Digestion of the PCR product and the pET 28b vector with the NdeI and BamHI enzymes (New England Biolabs) were performed, and the products isolated by agarose gel extraction. This allowed for subsequent ligation of the two fragments with T4 DNA ligase (New England Biolabs) in the presence of ATP. After ligation, the reactions were transformed into JM109 *E. coli* cells and plated on kanamycin supplemented LB agar plates. Colonies were picked, grown, and analyzed for correct

insertion of the INGAPrP gene. Ultimate verification of pINGAPrP was performed by DNA sequencing from the T7 forward primer. The DNA and protein sequences of pINGAPrP are shown in Figure 4.1.

```
DNA       AAGGAGA TAT ACC ATG GGC AGC AGC CAT CAT CAT CAT CAT CAC AGC
pINGAPrP   *RBS*          M   G   S   S   H   H   H   H   H   H   S

DNA       AGC GGC CTG GTG CCG CGC GGC AGC CAT ATG GAA CAA TCC CAG AAA
pINGAPrP   S   G   L   V   P   R   G   S   H   M   E   Q   S   Q   K

DNA       AAA CTG TCT TCT CCA CGC ATC AGC TGT CCC CAA GAA GCC CAA GCT
pINGAPrP   K   L   S   S   P   R   I   S   C   P   Q   E   A   Q   A

DNA       TAT GGC TCC TAT TGC TAT TTA CTG ATT CTG GAA CCA CAG ACC TGG
pINGAPrP  Y   G   S   Y   C   Y   L   L   I   L   E   P   Q   T   W

DNA       GCT AAT GCA GAG ATC CAC TGC CAG AAG CAT TTC TCA GGA CAC CTG
pINGAPrP  A   N   A   E   I   H   C   Q   K   H   F   S   G   H   L

DNA       GCA TTT CTG CTC ACT TAT GGT GAA ATT ATC TTT GTG TCC TCT CTG
pINGAPrP   A   F   L   L   T   Y   G   E   I   I   F   V   S   S   L

DNA       GTG AAA AAC AGT TTG ACC ACA TTC CCA TAC ATC TGG ATT GGA CTC
pINGAPrP   V   K   N   S   L   T   T   F   P   Y   I   W   I   G   L

DNA       CAT GAT CTG TCA CTT GGG AGT TTG CCC AAT GAA AAT GGA TGG AAG
pINGAPrP   H   D   L   S   L   G   S   L   P   N   E   N   G   W   K

DNA       TGG AGC AGC TCT GAC CCT CTG ACC TTC TAT AAC TGG GAG ATA CCA
pINGAPrP   W   S   S   S   D   P   L   T   F   Y   N   W   E   I   P

DNA       CCC TCC ATG TCT GCA CAC CAC GGC TAC TGT GCA GCT TTG TCT CAG
pINGPrP    P   S   M   S   A   H   H   G   Y   C   A   A   L   S   Q

DNA       GCC TCA GGT TAT CAG AAG TGG AGA GAT TAT TAT TGT GAC ACA ATA
pINGAPrP   A   S   G   Y   Q   K   W   R   D   Y   Y   C   D   T   I

DNA       TTT CCC TAT GTC TGC AAA TTC AAG GGT TAG
pINGAPrP   F   P   Y   V   C   K   F   K   G *STOP*
```

Figure 4.1 DNA and protein sequences for the pINGAPrP construct. RBS is the ribosome binding site.

4.3 Recombinant Expression of INGAPrP

Several *E. coli* cell lines were tried before an acceptable level of expression was achieved from Rosetta (DE3) cells (Novagen). These cells contain the pRARE plasmid

that codes for tRNAs rarely used by *E. coli* (Zhang et al., 1991). Rosetta (DE3) cells have been shown to increase eukaryotic gene expression levels in *E. coli* (Bukhtiyarova et al., 2004; Diallo et al., 2005; Hedayati et al., 2005). The pINGAPrP construct contains nine rare codons (Table 4.1). It is believed that these rare codons were causing expression difficulties in the other cell lines available for T7 expression systems, including DE3, DE3 (pLysS), and AD494. Several other constructs were prepared and tested in the various cell lines without pRARE, including a GST fusion protein and a pel-B fusion protein, with no significant amount of expression. A T5 expression system was also tried (The QIAexpressionist, 2001), in which the same INGAPrP gene, without the signal sequence, was cloned into a T5-compatible expression plasmid, pQE-31. Expression was attempted in the T5 expression cell lines M15 (pREP4) and SG13009 (pREP4) with no success.

Table 4.1 Rare codons found in pINGAPrP. The codons rarely used by *E. coli* are shown on the left, with the corresponding number of the codons found in INGAPrP on the right.

AGG	0
AGA	1
AUA	2
CUA	0
CGA	0
CGG	0
CCC	4
UCG	0
GGA	2

Expressions were routinely carried out in one liter of LB broth supplemented with kanamycin and chloramphenicol. The kanamycin selected for the pINGAPrP vector, while the chloramphenicol selected for the pRARE plasmid in the Rosetta (DE3) cells. Expressions were always inoculated with a colony from a fresh transformation plate ensuring the viability of the cells.

Optimal expression conditions were worked out for the pINGAPrP construct in the Rosetta (DE3) cells. An optical density, A_{600}, of 0.55 was found to be the best for induction. A final concentration of 1 mM IPTG was determined to give the highest expression levels of INGAPrP. It was also discovered that a long expression period at a low temperature was much better than the standard 3 to 4 hour expression at 37°C. Therefore, cells were grown at 37°C until the desired density was reached, then they were induced and placed at 27.7°C overnight. The overnight expressions were allowed to go for 15 to 17 hours.

In order to uniformly label the protein with ^{15}N for heteronuclear NMR experiments, we needed to use a different media than the LB. LB contains many sources of nitrogen and none of them are ^{15}N-labeled. Our first option was to use standard M-9 media. M-9 consists of salts (Na_2HPO_4, KH_2PO_4, NaCl, $MgSO_4$, and $CaCl_2$), appropriate antibiotics (kanamycin and chloramphenicol), glucose as a carbon source, and ^{15}N-labeled NH_4Cl.

M-9 expression was attempted with the same conditions as the LB, but with very poor yields (~0.3 mg/L cells). It is likely that the cells could not grow as well in the minimal media as they did in the LB. A new method was needed. It was found that there are pre-made media that contain only ^{15}N nitrogen sources. These nitrogen sources are

included in proteins, peptides, and vitamins, which will significantly aid the cellular growth and subsequent recombinant protein production.

The media chosen was Spectra-9 media (Spectra Stable Isotopes). The conditions worked-out with the LB media worked well with the Spectra-9. Yields were at least seven-fold higher than with the M-9 media. Although the yields from Spectra-9 were not as high as for the LB, enough protein was produced for heteronuclear NMR experiments.

4.4 Initial Purification (IMAC) of INGAPrP

The purification and refolding scheme I planned to follow was the same that had been used in our lab previously, to prepare the HTLV-1 protease. The idea is simply to purify the protein under denaturing conditions via an immobilized metal affinity chromatography (IMAC) column. This works by the interaction of Ni^{2+} and the hexahistidine tag at the N-terminus of the recombinant protein. After purification, the protein is refolded by simple dialysis into the desired buffer for further experiments.

IMAC purification for INGAPrP removed a large amount of contaminating proteins, but homogenous purification was never achieved. Imidazole gradients, pH gradients, and combinations of imidazole and pH were all attempted, but were unsuccessful in complete purification of INGAPrP.

The procedure for producing partially purified INGAPrP from IMAC followed a time-tested method for gathering protein from *E. coli*. Briefly, the protein from the Rosetta (DE3) cells was extracted from inclusion bodies with 6 M guanidine HCl and buffer A (20 mM Tris, 500 mM NaCl, 5 mM imidazole at pH 7.9). Guanidine HCl was empirically determined to yield higher protein recovery than urea. Once resuspended, the

cells were lysed by sonication while on ice. After two hours on ice, the lysate was centrifuged, producing a pellet and a supernatant. The pellet contained the cellular debris, while the supernatant contained all the proteins, in guanidine.

The IMAC used was Ni-NTA (QIAGEN) and was prepared by flowing several column volumes of 100 mM $NiSO_4$ over the resin. As mentioned, several purification schemes were tried. All of the purification procedures were carried out under denaturing conditions in 6 M urea. The SDS-PAGE analysis for two of the schemes are shown below. Figure 4.2 is representative of a gradual imidazole gradient used to wash away impurities before elution. Figure 4.3 shows the results from a pH gradient to wash away contaminants, followed by an imidazole elution. It is evident that neither of these methods produce homogenous INGAPrP.

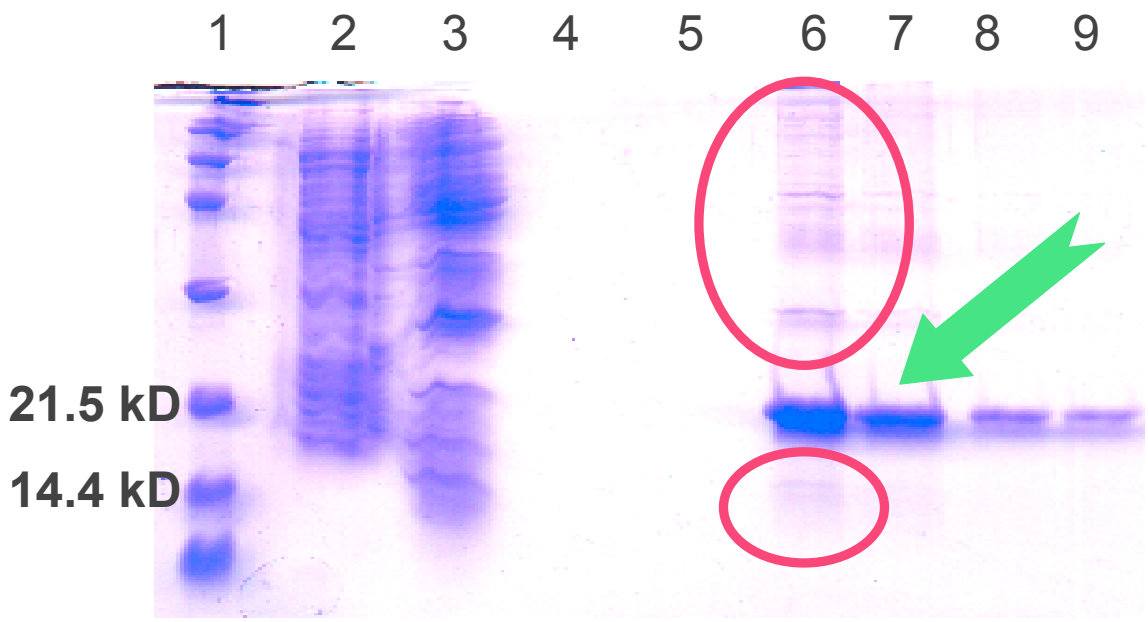

Figure 4.2 Analysis of IMAC purification via imidazole gradient. Lane 1: peptide ladder (21.5 kD and 14.4 kD markers are labeled); Lanes 2-5: imidazole wash fractions (5, 20, 50, 75 mM imidazole, respectively); Lanes 6-9: elution fractions (300 mM imidazole). The red circles high-light the impurities in the elution fractions. The green arrow shows INGAPrP at ~19.5 kD.

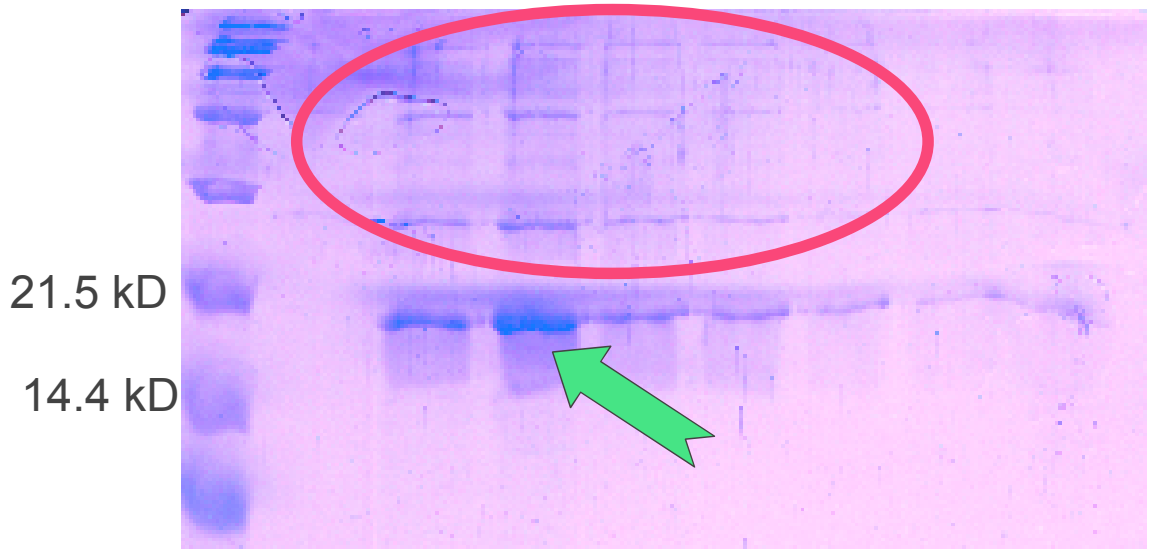

Figure 4.3 Analysis of IMAC purification via pH gradient. The left-most lane is the peptide ladder (21.5 kD and 14.4 kD markers are labeled). The other lanes are all elution fractions following a pH gradient of 7.5, 6.3, 5.9, and 4.5. The red circle high-lights the impurities following the pH gradient. The green arrow indicates INGAPrP at ~19.5 kD.

The method used to produce INGAPrP for further purification utilized an imidazole gradient of 5 mM, 20 mM, and then elution with 1 M imidazole. The gradient was carried out under denaturing conditions in 6 M urea. The SDS-PAGE analysis of this protocol can be seen in Figure 4.4.

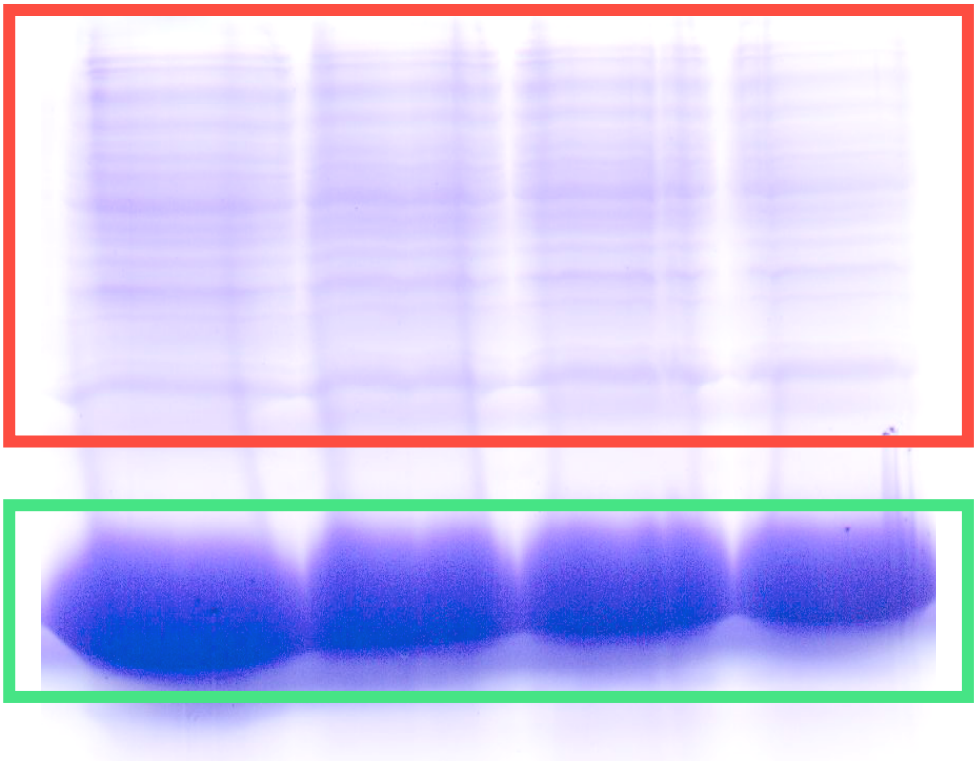

Figure 4.4 SDS-PAGE analysis of the elution fractions from IMAC following washes with 5 mM and 20 mM imidazole. INGAPrP is shown in the green box. The impurities are shown in the red box.

Since the protein was not purified to a high enough level, a second purification step was needed. It was initially planned to dialyze the elution fractions from IMAC into a native buffer, then attempt size-exclusion chromatography or ion-exchange chromatography. This way the protein could be refolded, and then purified. However, this proved unsuccessful because the protein would not stay in solution upon dialysis. Numerous dialysis methods were tried using literature methods (Clark, 1998). The first attempt was to dialyze directly from the 6 M urea in the elution fractions into a large excess of native buffer. Several conditions for the native buffers were altered such as pH, salt and buffer composition, and temperature, with no change. A step-wise urea gradient

was attempted in which the protein was first dialyzed against a 5 M urea buffer, then a 4 M, and 3 M. However, protein continually precipitated at a urea concentration lower than ~3.5 M. Several stabilizing additives were also tried, such as arginine, whose structure mimics that of guanidine, and glycerol, which acts as a way to form small micelles around the protein.

Ultimately, it was decided to attempt purification and folding on the same column. This would allow us to proceed with the partially pure, denatured, INGAPrP from the IMAC step and move onto an FPLC system. It was believed that we could purify the protein and fold it all on a gel-permeation column (Gu et al., 2001; Shi et al., 2003). To try this, we concentrated the protein from IMAC down to a small volume for FPLC work.

4.5 Secondary Purification (GPC) and Folding of INGAPrP

Gel permeation chromatography (GPC), also known as size-exclusion chromatography, separates proteins based on their molecular weights. The retention times of the proteins are inversely related to their molecular weights. Large proteins come off the column first and small proteins elute later. This is due to the small proteins interacting with the beads longer than the large proteins. The beads are very porous, which allows for smaller proteins to pass into the beads whereas large proteins can not. GPC was our method of choice to further purify the recombinantly expressed INGAPrP.

A Superdex 75 HR 10/30 column (Pharmacia) was used. This column was in-line on an Akta Explorer FPLC system. The FPLC system is located in a cold room allowing

for all purification steps to occur at 4°C. This minimizes sample degradation and bacteria or mold contamination in buffers.

To purify the protein, the column was first equilibrated with the buffer to be used for purification, which was a 25 mM Tris, 150 mM NaCl buffer at pH 8.5 (at 4°C). The NaCl was added because the column specs call for at least 150 mM NaCl when purifying protein. The Tris buffer at a pH of 8.5 was chosen because this was the pH at which INGAP was refolded (Vinik et al., 1998).

After equilibration, small protein samples, 250-500 μL, were injected onto the column. The Tris/NaCl buffer was used as the mobile phase. Separation of the proteins was evident by observing traces of A_{280} versus elution volume. A typical trace from the GPC can be seen in Figure 4.5.

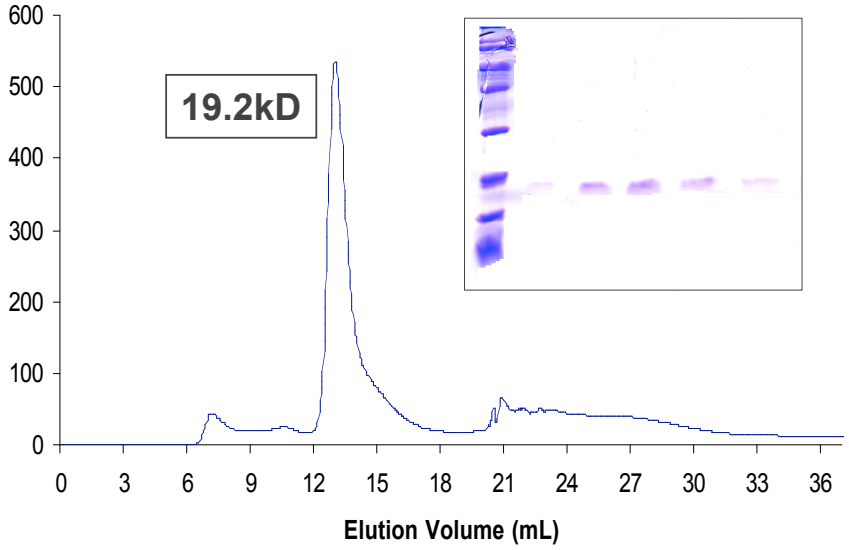

Figure 4.5 A typical chromatogram from GPC purification of INGAPrP. INGAPrP elutes from 12.5 mL to 14.5 mL and is the largest peak. The retention time corresponds to a molecular weight of ~19.2 kD. The inset shows a typical SDS-PAGE gel for the INGAPrP fractions following GPC purification. No impurities are observed.

The largest peak from the GPC purification was found to be INGAPrP. A standard curve for the Superdex 75 HR column was worked out using proteins of known molecular weight by Dr. Charlie Oldham. An equation was generated to determine the molecular weight of an unknown protein based on its retention time on the column. The protein with a retention time of ~12.8 mL (retention time at maximal A_{280}) was a protein with an approximate molecular weight of 19.2 kD. This was in the range of that expected for INGAPrP. The largest peak was run on an SDS-PAGE gel and found to match-up with the most abundant protein from the IMAC column elution fraction (Figure 4.6). This was identified as INGAPrP.

Figure 4.6 SDS-PAGE analysis of the largest peak from GPC (left lane) and elution fraction from IMAC (right lane). The peak from GPC matches the molecular weight of the most abundant protein from the IMAC elutions.

MALDI analysis showed that the largest peak from the GPC column had a molecular weight of 19.404 kD (Figure 4.7). This is 80 daltons more than the predicted molecular weight. This is likely the result of adducts, such as sodium, or the matrix, being bound to the INGAPrP protein.

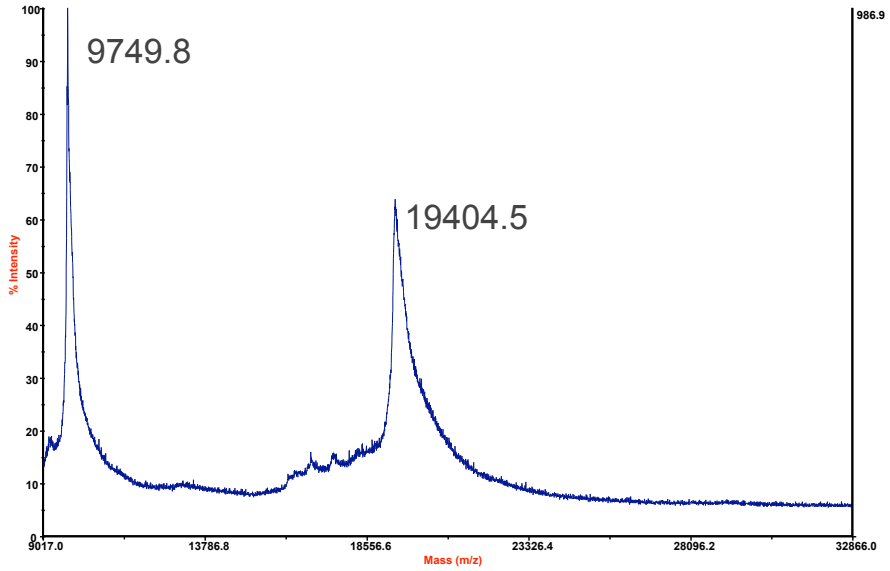

Figure 4.7 MALDI analysis of purified INGAPrP. The parent ion (19.404 kD) gives the molecular weight of INGAPrP. The Z = 2+ peak is also present at 9.749 kD.

^{15}N-labeled protein was purified in the same manner as for non-labeled samples. Final yields after the GPC purification were 3.5 mg/L-cells from LB media, 0.3 mg/L-cells from M-9 media, and 2.0 mg/L from Spectra-9 media.

To test the hypothesis that the protein from the GPC would be folded, we needed to run a ^{1}H-^{15}N HSQC NMR experiment and CD experiments. These are reported in the following chapter.

4.6 Conclusions

The pINGAPrP construct is capable of expressing mg quantities of INGAPrP in Rosetta (DE3) *E. coli* cells. This method will allow for either ^{15}N-labeled or unlabeled samples of protein to be produced. Purification of INGAPrP by IMAC is insufficient to yield highly pure protein. Therefore a second purification step, GPC, was needed. This protocol yields highly pure INGAPrP that can be used for subsequent structural determination studies.

CHAPTER 5

NMR, CD, AND OTHER STUDIES OF INGAPrP

5.1 NMR Experiments

The goal of this project was to determine the three-dimensional solution structure of INGAPrP. After isolation of the protein, the next step was to begin NMR experimentation. After GPC purification, the INGAPrP fractions were concentrated down to a useful volume for NMR and D_2O was added to a final concentration of 10%.

The one-dimensional experiment was conducted and concluded to be a success. The purpose of the one-dimensional ^1H NMR was to look for characteristics of a folded protein. There are two important regions for this. The first is the amide region, or the 7-11 ppm region. A folded protein will have good resolution in this region with many sharp peaks (Figure 5.1). The high resolution indicates that the amide protons have their own unique chemical environment. An unfolded protein will have poor resolution in the amide region, and will have very few, if any, sharp peaks (Figure 5.1).

The second important region in the ^1H NMR spectrum is the area below zero ppm. A folded protein will usually have a characteristic peak, or peaks, in this region (Figure 5.1). In an unfolded protein, the peaks below zero ppm will be conspicuously absent (Figure 5.1). These peaks arise from a shielding effect experienced by a given proton. The shielding is caused by aromatic side-chains interacting with other protons in three-dimensional space, indicative of a folded structure.

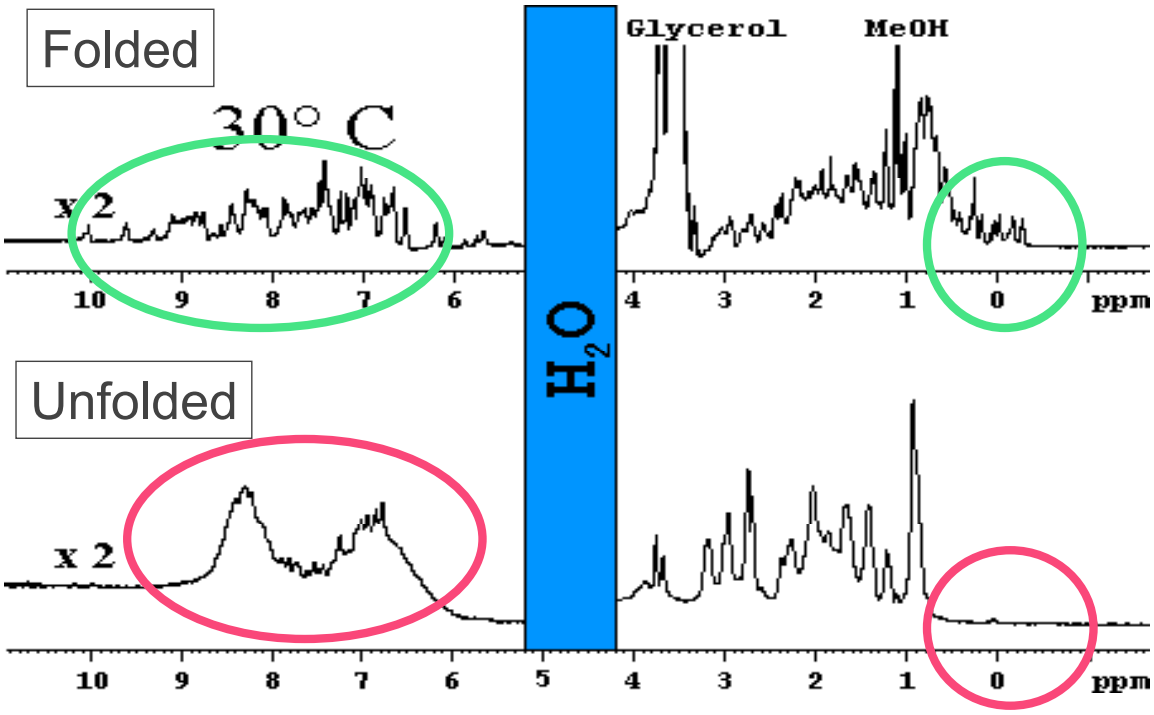

Figure 5.1 One-dimensional ^{1}H NMR comparing "folded" and "unfolded" protein samples. The top spectrum illustrates an example of a folded protein. The bottom spectrum is for an unfolded sample. The areas of interest are shown inside the colored circles. Note that neither of these examples are of INGAPrP, but rather are used to illustrate "typical" one-dimensional NMR characteristics of proteins.

The INGAPrP sample, prepared from the IMAC and GPC chromatography, displayed the characteristics of a folded protein in the one-dimensional ^{1}H NMR. The spectrum can be seen in Figure 5.2. With this initial promising result, the next step was to perform a two-dimensional experiment.

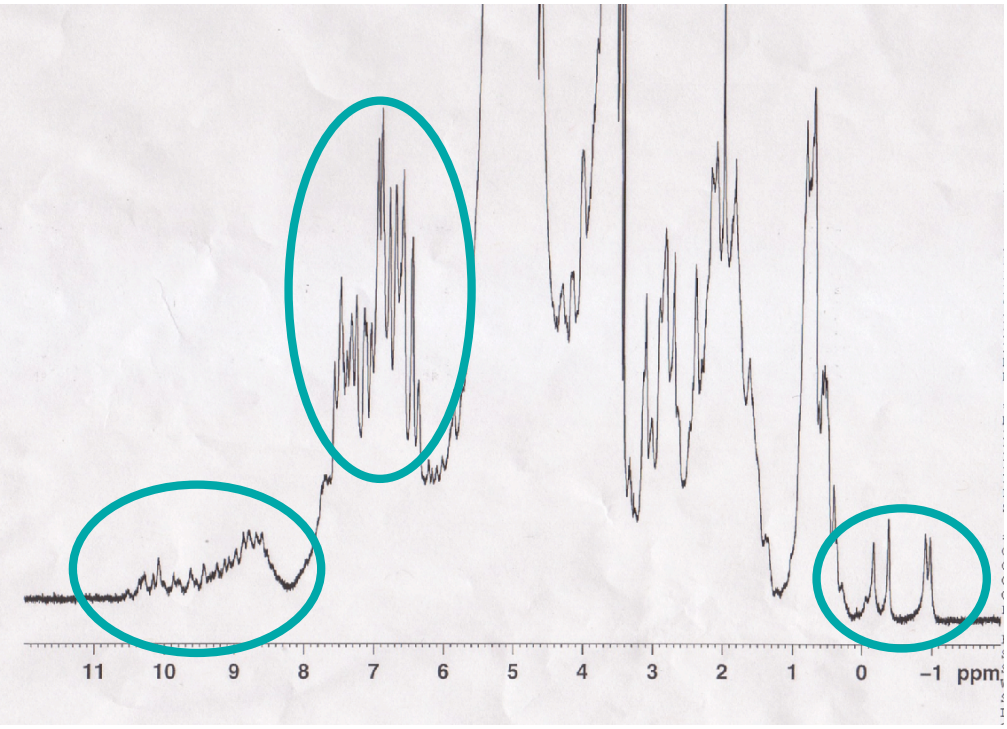

Figure 5.2 One-dimensional ^{1}H NMR of purified INGAPrP. The areas of interest are shown inside the blue circles. The protein appears to be folded.

The two-dimensional experiment of choice is the ^{1}H-^{15}N HSQC. The HSQC is chosen for several reasons. First, it will give much better resolution of the amide region than the ^{1}H NMR because of its two dimensions. Second, it is relatively easy to prepare a sample for the experiment by labeling it with ^{15}N. This labeling is also much cheaper than other labeling choices, such as ^{13}C. Also, the HSQC is a fairly quick experiment. This experiment is the first step in multi-dimensional NMR experiments for structure determination.

The ^{1}H-^{15}N HSQC examines the protons connected to nitrogens. These are mostly from the protein's backbone, with some coming from the various nitrogen-containing side-chains. A folded protein will give a broad dispersion of peaks in the HSQC, whereas a non-folded protein will have poor resolution in the HSQC (Figure 5.3).

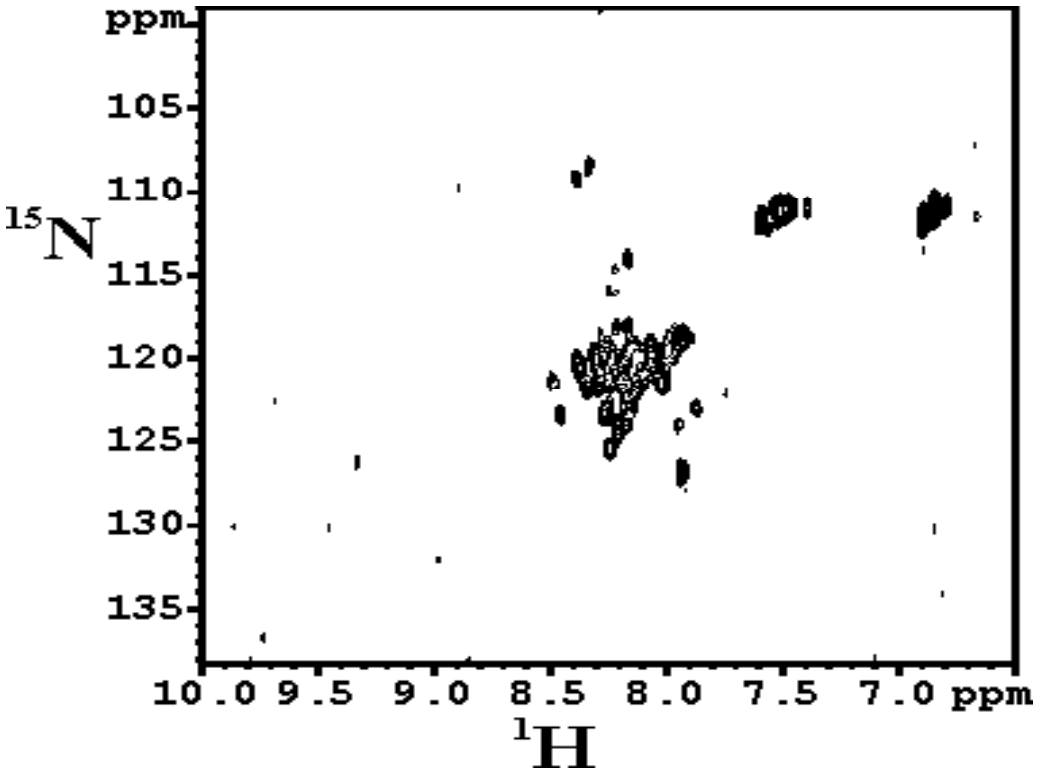

Figure 5.3 ^{1}H-^{15}N HSQC of an unfolded protein. This is an example illustrating the poor resolution of an unfolded protein, and is not INGAPrP.

The INGAPrP sample was concentrated further and placed into a Shigemi tube for the ^{1}H-^{15}N HSQC. This was done to increase the effective concentration of the protein to obtain an ^{1}H-^{15}N HSQC in a shorter time. The resulting spectrum can be seen in Figure

5.4. It is apparent that the protein is folded because of the excellent dispersion of the peaks.

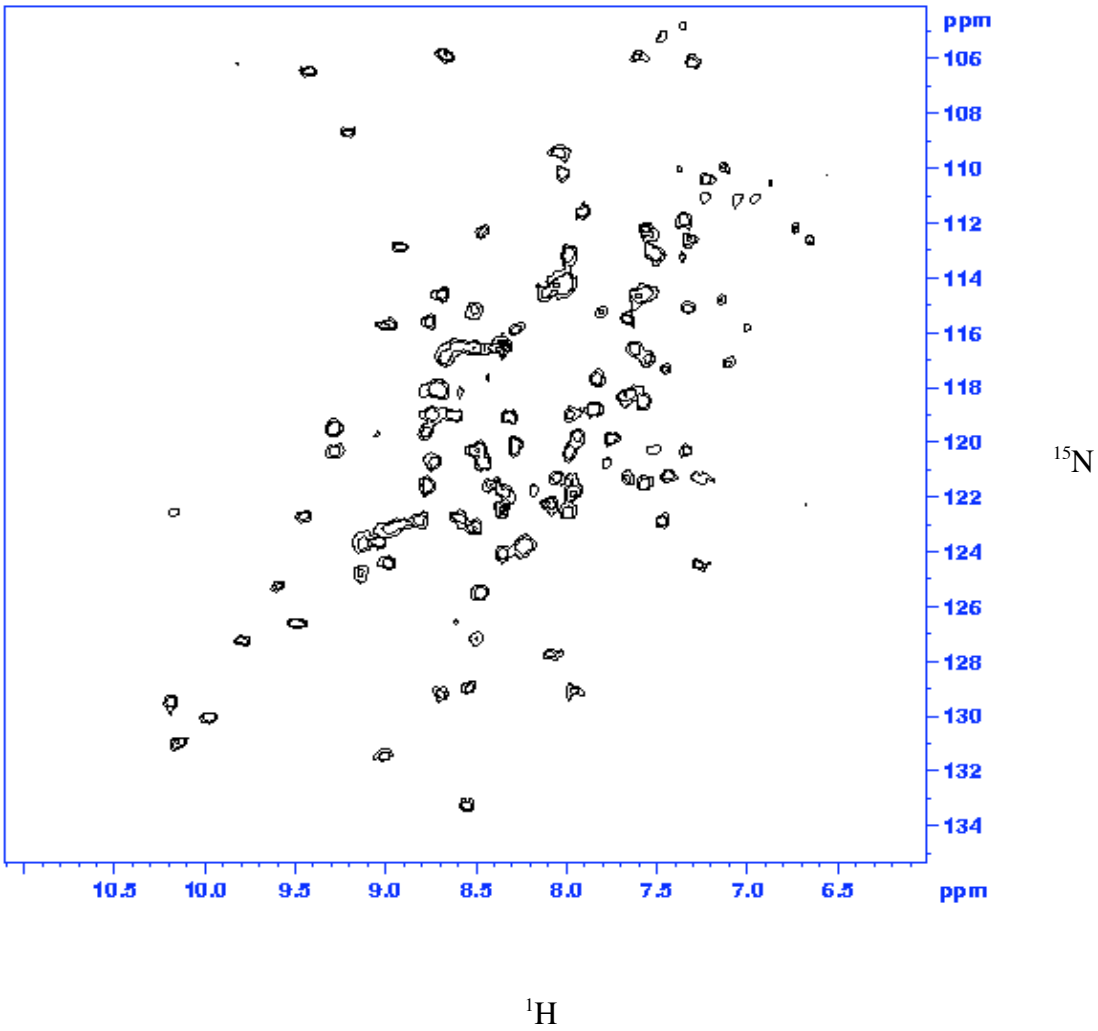

Figure 5.4 ^{1}H-^{15}N HSQC of purified INGAPrP. The signal dispersion is indicative of a folded state for INGAPrP.

This is a very promising result indicating that our method of purification also yields a folded protein. The protocol can now be used to determine the three-dimensional

structure of INGAPrP. It should be noted that although the protein is folded, there is no way to verify that this fold is the "correct", or native, fold of INGAPrP. The way to test this would be to isolate INGAPrP directly from mouse pancreata and compare the HSQCs and CD spectra. Another option would be to test the recombinant INGAPrP for activity in a functional assay. If the protein is active then it would be considered to have the correct fold. An assay is underway with Dr. Alberto Hayek's group at the University of California at San Diego, however, results from the assay are unavailable at the time of preparing this document.

5.2 CD

Circular dichroism (CD) is an experiment that is used to observe secondary structure in proteins. Circular dichroism is the result of a molecule absorbing left and right circularly polarized light to a different extent. The amide chromophore of the peptide bond dominates the CD spectra below 250 nm. Secondary structural elements in a protein will cause characteristic CD spectra. Examples of structural elements are shown in Figure 5.5.

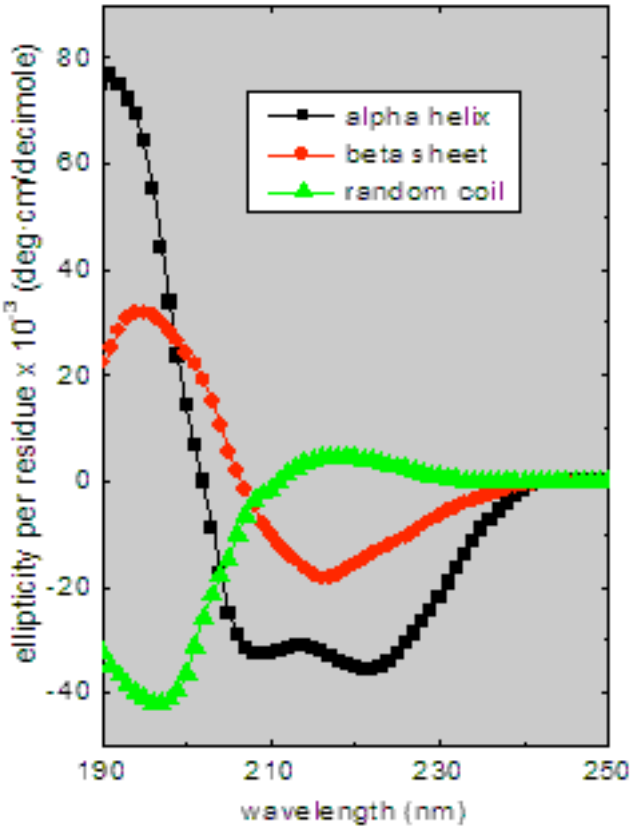

Figure 5.5 CD spectra showing characteristics of alpha-helical, beta-sheet, and random coil structures in proteins.

A CD was run on INGAPrP to complement the NMR conclusion that the purified INGAPrP was folded. The buffer and salt concentrations were reduced to 5 mM Tris and 30 mM NaCl to keep the voltage of the CD within an accurate range. The pH was kept at 8.5. The sample was diluted to a concentration of 25 □g/mL.

The CD results can be seen in Figure 5.6. From visual inspection, it appeared that there was some helical content in the spectrum. This was verified by deconvolusion of the CD data via a program called SELCON3 (Sreerama and Woody, 1993). SELCON3

works by taking CD data from proteins with known structures and then compares this data set with the input CD data. This program concluded that there were several helices and at least one beta-sheet in INGAPrP from its CD.

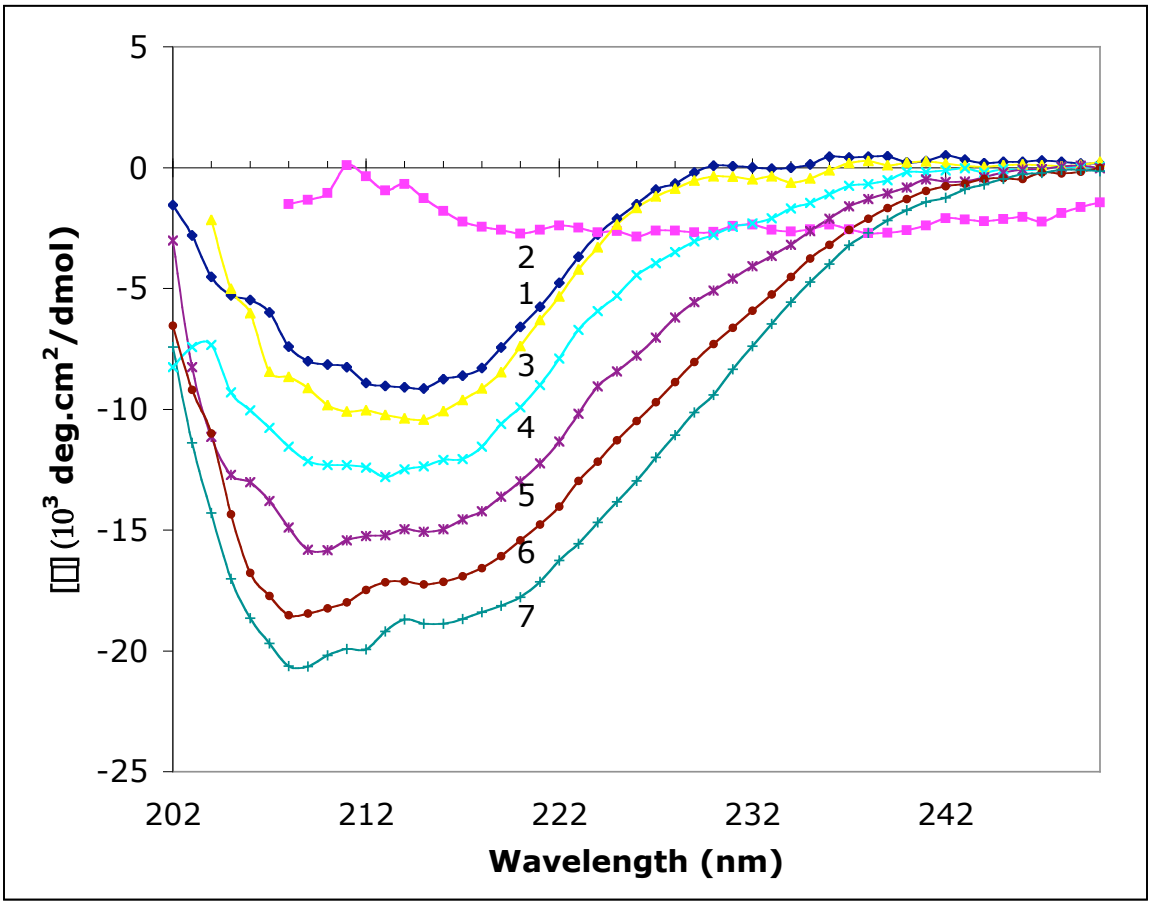

Figure 5.6 CD spectra of 25 μg/mL INGAPrP under various conditions. All TFE measurements were done at 22.0°C. 1: 5 mM Tris, 30 mM NaCl, pH 8.5 at 22.0°C. 2: 5mM Tris, 30 mM NaCl, pH 8.5 at 90.0°C. 3: 20% TFE. 4: 30% TFE. 5: 40% TFE. 6: 50% TFE. 7: 60% TFE.

Further verification that there is helical content in INGAPrP was examined by titrating in TFE to the protein sample. TFE is known to induce helical content in proteins. It is thought to work by strengthening hydrogen bonding networks, which are

crucial in helices. A range of 20-60% TFE was used on INGAPrP. With increasing TFE concentration it is apparent that the helical content is increasing (Figure 5.6).

Finally, to further establish the fact that INGAPrP was folded, a simple denaturation experiment was performed and a CD acquired. INGAPrP was denatured by heating it at 90.0°C for 15 minutes. The denatured INGAPrP showed no signs of secondary structure (Figure 5.6).

The CD data demonstrates the presence of secondary structure within INGAPrP. It also indicates the presence of helical content within INGAPrP, which is useful information for further structural studies of the protein. Thus, the CD data complements the NMR data indicating that INGAPrP is folded, and contains secondary structural elements.

5.3 Degradation of INGAPrP

INGAPrP was found to degrade at 4°C after approximately seven days following GPC purification. This was observed by SDS-PAGE analysis and can be seen in Figure 5.7. It was determined that the protein degraded to 16.1 kD by MALDI analysis (Figure 5.8). This meant that the protein was "losing" roughly 30 amino acids from the full-length protein expressed from the pINGAPrP construct. The two possibilities for the 16.1 kD protein can be seen in Figure 5.9.

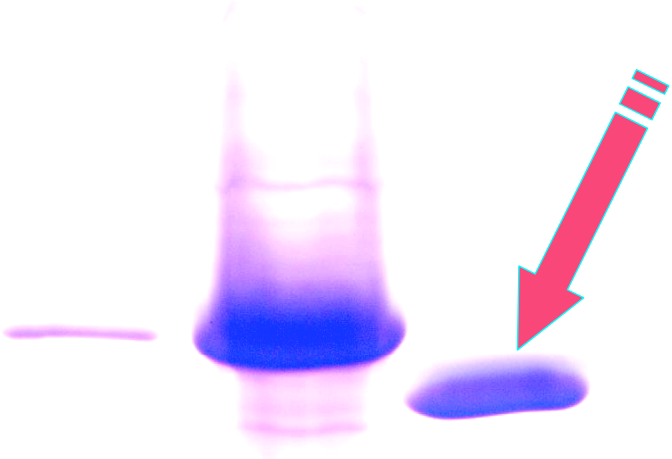

Figure 5.7 SDS-PAGE analysis showing INGAPrP degradation after several days at 4°C. The left lane shows INGAPrP following GPC. The center lane shows INGAPrP following IMAC purification. The right lane shows degraded INGAPrP (red arrow).

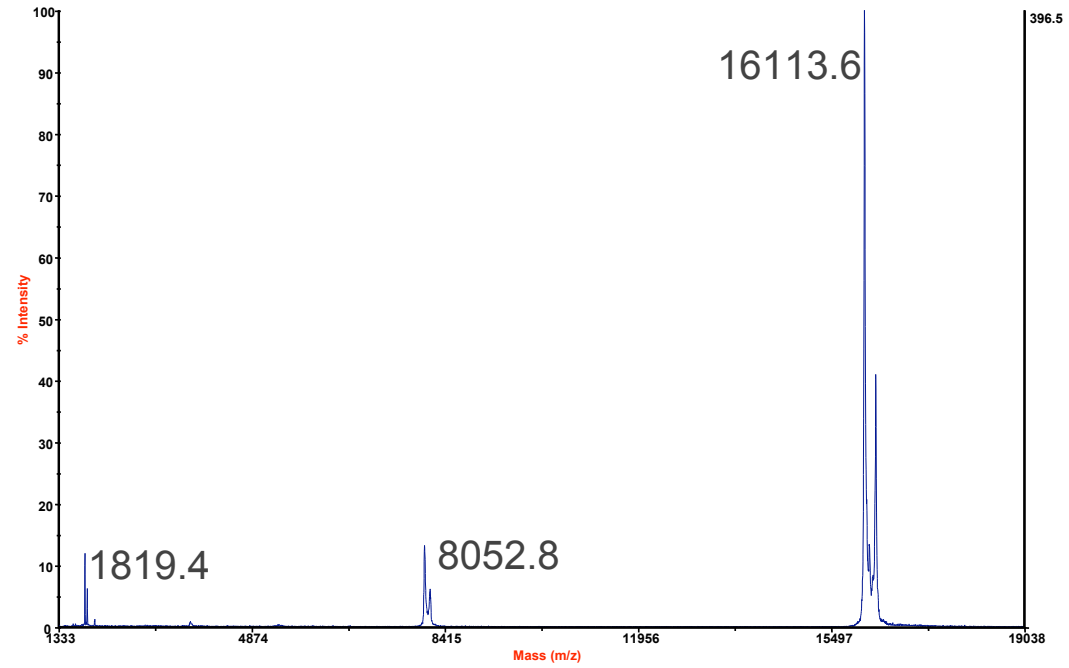

Figure 5.8 MALDI analysis of degraded INGAPrP. The parent peak is 16.113 kD and the Z = 2+ peak is 8.052 kD.

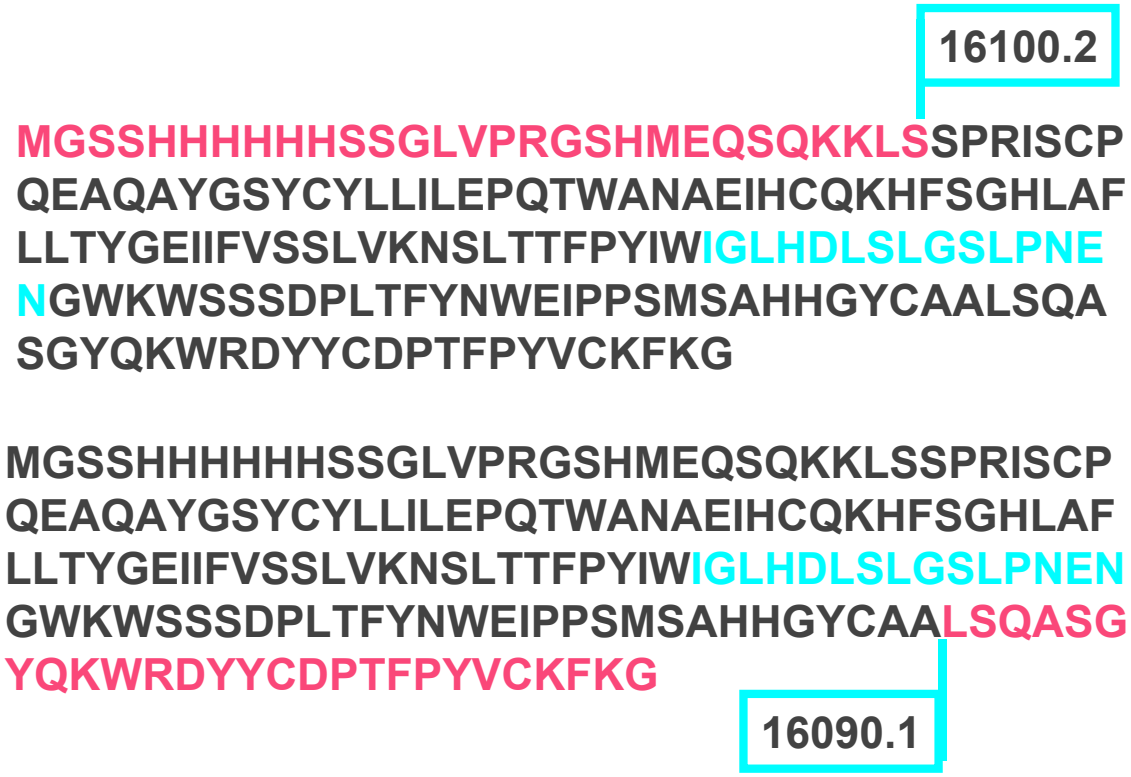

Figure 5.9 Two possibilities for the 16.1 kD fragment of INGAPrP. The residues that would be removed following degradation are shown in red, with the corresponding molecular weight of the remaining protein shown in the blue box (in kD). The putative pentadecapeptide region is shown in blue.

In order to determine which of the two possibilities in Figure 5.9 is correct, the degraded sample was submitted for Edman degradation. The sample was brought to Dr. Jan Pohl at the Emory University Proteomics Facility. Edman degradation determines the N-terminal sequence of a protein. The protein was subjected to eight rounds of Edman sequencing and the results were SPRISCPQ as the N-terminus. This means that the protein is being degraded at the N-terminus of the full-length protein from the pINGAPrP construct (top sequence in Figure 5.9). Also, this result matches exactly with the MALDI data indicating the degraded protein has a molecular weight of 16.1 kD.

The most logical explanation for the degradation is that a protease is cleaving the INGAPrP. Trace amounts of protease could theoretically remain with INGAPrP following the GPC purification. However, since cleavage is occurring between two serine residues (Ser29 and Ser30), it is hard to say why the hypothetical protease would not cleave the other Ser-Ser junctions within the remaining protein (Ser84-Ser85, Ser118-Ser119, and Ser119-Ser120). One possibility is that the other Ser-Ser sites are buried within the protein and are not accessible to the protease. It could also be reasoned that the residues surrounding the other Ser-Ser sites do not accommodate the protease's active site, and thus can not be cleaved. It should be pointed out that there is another Ser-Ser site (Ser11-Ser12) within the N-terminus, which appears to be cleaved as evidenced by SDS-PAGE (Figure 5.10). This apparent specificity adds to the hypothesis that a protease is responsible for the degradation at INGAPrP's N-terminus.

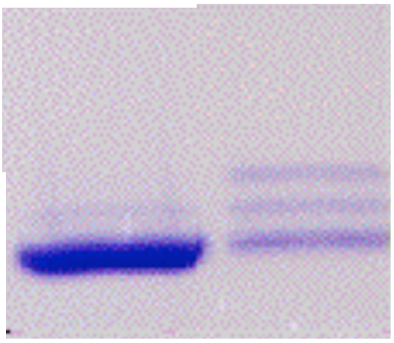

Figure 5.10 SDS-PAGE showing a potential secondary degradation site within INGAPrP's N-terminus. The left lane shows a sample that has degraded almost completely to 16.1 kD. The right lane shows a sample that has three bands; the top is 19.4 kD INGAPrP, the middle is an apparent fragment produced from cleavage between Ser11-Ser12, and the lowest is the 16.1 kD INGAPrP.

To test the hypothesis that INGAPrP is being degraded by a protease, a simple experiment was designed. A 19.4 kD INGAPrP sample was thawed from -20°C and supplemented with a protease inhibitor cocktail, P8215, from Sigma-Aldrich. The cocktail contained AEBSF, E-64, Pepstatin A, and 1,10-Phenanthroline. These molecules inhibit serine proteases, cysteine proteases, aspartic acid proteases, and metalloproteases, respectively. After addition of the cocktail to the protein sample, the sample was placed at 4°C for three weeks with aliquots being taken daily to monitor degradation. The positive control was a 19.4 kD INGAPrP sample thawed from -20°C supplemented with no protease inhibitors.

The results were not expected. The positive control showed no signs of degradation after three weeks at 4°C (Figure 5.11). The sample with the inhibitor cocktail also showed no degradation following three weeks at 4°C (Figure 5.11). Since the positive control did not work, these test results can not verify the idea that the degradation is being caused by a protease.

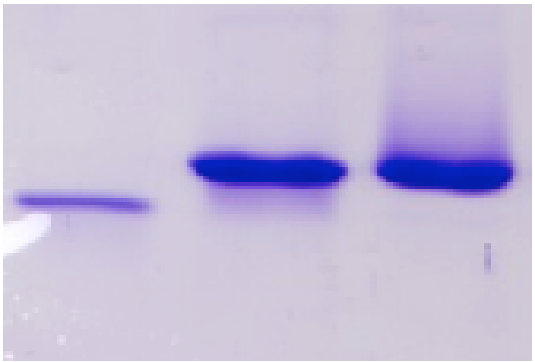

Figure 5.11 SDS-PAGE analysis of anti-degradation experiment. The left lane shows a 16.1 kD INGAPrP sample. The center lane shows a positive-control sample after three weeks at 4°C. The right lane shows a sample with the protease inhibitor cocktail after three weeks at 4°C.

It is unknown why the positive control did not undergo any degradation. The only difference between this full-length sample and other full-length samples that have shown degradation is that this sample was frozen at -20°C for several weeks. Samples form the same batch of GPC purified protein underwent subsequent degradation, possibly because they were never frozen, but rather kept at 4°C immediately following purification. This leads to the conclusion that whatever is causing the degradation is somehow altered by freezing at -20°C, and thus rendered inactive. This may be a possible way to prevent degradation in the future. However, before this method of freezing INGAPrP following GPC purification is used, it should first be demonstrated that there is no alteration to the protein fold following freezing. This could be monitored with NMR or CD.

In conclusion, regardless of how, or why, the protein is cleaved, the degradation should not significantly affect the NMR structure determination. Only eight residues from the INGAPrP protein are lost, with the other 21 being from the pET 28b vector (Figure 5.9). Also, the putative INGAPrP pentadecapeptide region is still intact (Figure 5.9). This is the region of ultimate interest for our purposes.

5.4 Homology Modeling

A homology model was constructed for reference purposes. A modeling program called molecular operating environment (MOE) was used. A homology model was constructed using the x-ray crystal structure of Lithostathine (Reg), PDB entry 1QDD, as the template. Lithostathine has 42% identity to INGAPrP and thus serves as a reasonable template for INGAPrP modeling.

The homology model can be seen in Figure 5.12. It indicates some helical content and some beta-sheet content, consistent with the CD data. In my opinion, the interesting part of the homology model is the area of the INGAPrP pentadecapeptide region. This region is shown in black in Figure 5.12. It is apparent that the 15-mer is solvent exposed and is in a turn conformation. This could be a conformation that is able to bind and activate its receptor. This conformation will need to be verified experimentally by NMR, or x-ray, structure determination before it will be of use for rational drug design.

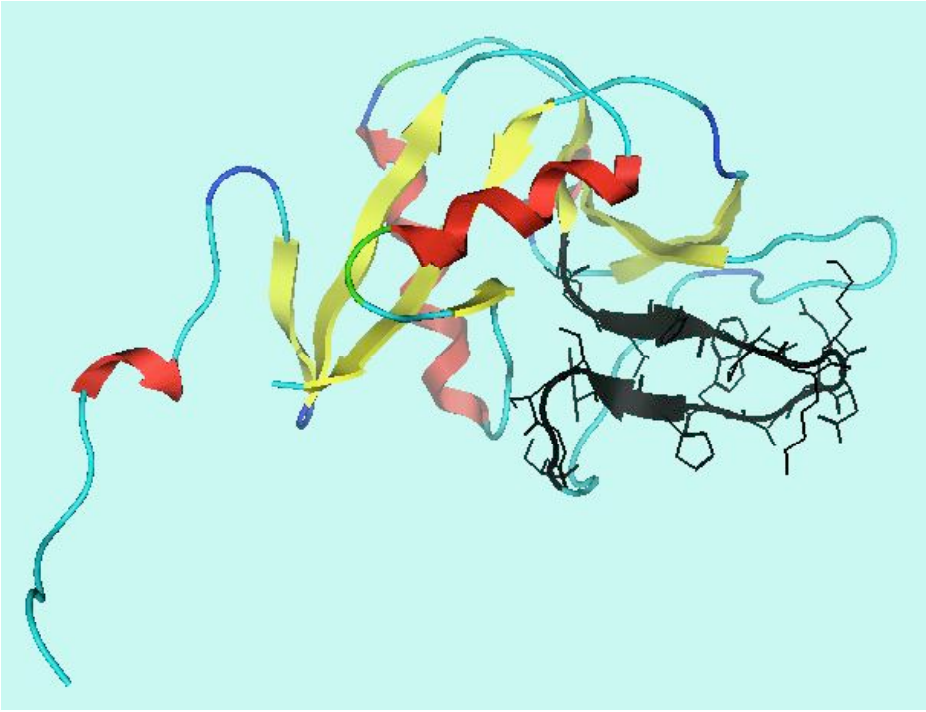

Figure 5.12 Homology model of INGAPrP. The model shows both helical and sheet contents. The pentadecapeptide region is shown in black.

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

This is the first report of the expression, purification, and refolding of INGAPrP. Also reported is the initial step in NMR structure determination of the protein, the ^{1}H-^{15}N HSQC. Circular dichroism experiments indicate that INGAPrP contains helical elements in its secondary structure. Although this is a start for our ultimate goals with this project, much is left to be done.

There are two logical steps to be taken following the work presented here. The first step is to fully determine the three-dimensional NMR solution structure of INGAPrP. The ^{1}H-^{15}N HSQC is proof that the protocol for isolation of INGAPrP yields a folded protein for the conditions tested, and therefore can be used for structure determination. The second main objective should be to investigate the activity of INGAPrP. A collaboration with Dr. Alberto Hayek at the University of California at San Diego is currently underway to test the activity of INGAPrP in proliferating and/or differentiating human ductal cells and/or human fetal pancreatic cells *in vitro*. This should provide an answer as to the desired, and anticipated activity of INGAPrP.

There are several other projects that can be imagined pertaining to the work described here. A mutant of INGAPrP was prepared by myself in which the putative INGAPrP pentadecapeptide region was successfully mutated to the INGAP pentadecapeptide (i.e. IGLHDLSLGSLPNEN to IGLHDPSHGTLPNGS). This would be an interesting molecule to look at structurally, because it should alleviate the problems we had with looking at the 15-mer in solution, since the INGAPrP protein would likely

act as a scaffold to hold the 15-mer in place. It would also be interesting to test the activity of this mutant.

Assuming that INGAPrP has activity similar to INGAP, it would be interesting to evaluate which part of the protein is the active portion. Studies should also be done with the putative INGAPrP pentadecapeptide *in vitro*, as well as the putative Reg pentadecapeptide, to compare the activity of these peptides with that of the INGAP 15-mer. Further studies along the same vein would be to perform an alanine-scan on the various peptides. For example, one could mutate each of the 15 residues in the INGAP pentadecapeptide to an alanine to produce 15 separate peptides. The activity of these peptides could indicate which residues are critical for activity and which ones could be altered. A subsequent study could then be envisioned in which a library of compounds is made by using various amino acids in the place of non-critical residues within the 15-mer.

The last set of experiments that are of interest would involve identification of a receptor for INGAPrP, INGAP, and the INGAP pentadecapeptide. There are some candidate receptors from the literature, including the rat Reg receptor and its human homologs. Various libraries of receptors could also be screened. Once a receptor is identified, it would be useful to investigate the signaling cascade that these differentiation factors induce. This could lead to additional therapeutic targets. A receptor would also be useful for SAR by NMR studies in which one could determine precisely which residues of the INGAPrP protein, or the INGAP pentadecapeptide, bind to the receptor (Shuker et al., 1996). This would provide key information for drug development of smaller molecules.

In summary, a solid foundation has been laid for the INGAPrP project, but much is left to be explored and discovered. It is my belief that once these additional projects are completed, a vast amount of information would be generated and could be utilized for improved drug molecules.

CHAPTER 7

MATERIALS AND METHODS

INGAPrP expression plasmid construct

INGAPrP cDNA (accession number **AA822059**) was obtained from ResGen. The INGAPrP gene without the putative signal peptide was produced via PCR with the primers 5'-CGTAGTCATATGGAACAATCCCAGAAA-3' and 5'-CGTAGTGGATCCCTAACCCTTGAATTT-3'. The underlined bases correspond to NdeI and BamHI restriction enzyme sites, respectively. The PCR was carried out over 30 cycles of 95°C for 1 minute, 58°C for 1 minute, and 72°C for 1 minute using Vent DNA polymerase (New England Biolabs). The expression plasmid pET 28b (Novagen) was used as the vector. Double enzymatic digests with NdeI and BamHI (New England Biolabs) were performed on both the PCR product and the pET 28b plasmid. A subsequent ligation reaction was performed with T4 DNA Ligase (New England Biolabs) in the presence of ATP to yield the desired construct, pINGAPrP. The ligation reaction was transformed into JM109 cells and grown on LB agar supplemented with 30 μg/mL kanamycin. A colony was picked from the agar plate and grown in LB media supplemented with kanamycin. A cell stock was made by adding 500 μL 7% DMSO to 500 μL cells and frozen at -80°C. DNA from the cells was isolated by Qiagen preparation. DNA sequencing, carried out at the Georgia Tech Biology Core Lab, using the T7 forward primer, verified the accuracy of pINGAPrP.

Expression of INGAPrP

pINGAPrP was transformed into Rosetta (DE3) cells (Novagen) and plated on LB-kanamycin/chloramphenicol (30 μg/ml and 34 μg/ml respectively). Colonies were picked and grown in flasks containing one-liter LB or Spectra-9 media (Spectra Stable Isotopes) supplemented with kanamycin and chloramphenicol. The growths were carried out at 37°C until an $O.D._{600}$ of 0.5-0.6 was reached, at which time IPTG was added to a final concentration of 1 mM. The cells were left in the presence of IPTG at 27.7°C for 15-20 hours to produce INGAPrP. Cells were pelleted at 5K RPM and resuspended in a denaturing buffer composed of buffer A (20 mM Tris, 500 mM NaCl, 5 mM imidazole at pH 7.9) with 6M guanidine HCl. The resuspended cells were lysed by sonication and left on ice for 3 hours to allow extraction of the protein in inclusion bodies. The mixture was spun at 14K RPM to separate the cellular debris from the protein solution. The supernatant containing the INGAPrP protein was saved for purification.

Initial Purification of INGAPrP by IMAC Chromatography

A 5 mL Ni-NTA (QIAGEN) resin bed was used for the IMAC purification step. All steps were carried out at 22°C. The column was equilibrated with 4 column volumes (CV) 0.5 M NaOH, 5 CV dI H_2O, 4 CV 100 mM $NiSO_4$, 3 CV buffer C (buffer A with 20 mM imidazole and 6 M urea) and then the supernatant was flowed over the resin. The protein was purified with 8 CV buffer B (buffer A with 6 M urea), 7 CV buffer C, and eluted with 6 CV buffer D (buffer B with 1 M imidazole). The elution fraction was concentrated down to a final volume of 5 mL for secondary purification. Protein was concentrated with Amacon concentrators with 3.5 kD MWCO.

Secondary Purification and Folding of INGAPrP by GPC Chromatography

A Superdex 75 HR column was used for GPC purification in-line on an AKTA Explorer FPLC system at 4°C. The column was equilibrated with 2 CV 25 mM Tris, 150 mM NaCl at pH 8.5. 500 μL protein from the IMAC column was injected onto the GPC column and the 25 mM Tris, 150 mM NaCl buffer was used as the mobile phase. The largest peak in the chromatogram, corresponding to a ~19.2 kD protein was saved at 4°C for NMR studies. Multiple injections of 500 μL protein were made and the INGAPrP fractions were pooled. MALDI analysis of the resultant protein was done at Georgia Tech by Dr. Cameron Sullards.

NMR Spectroscopy

A 0.5 mM protein sample was prepared by concentrating the GPC purified protein to a final volume of 300 μL after the pH of the solution was adjusted to 6.9 and D_2O was added to a final concentration of 10%. The protein solution was placed into a Shigemi tube and NMR data was collected on a Bruker AVANCE DRX 500 at 22°C. The water signal was suppressed with WATERGATE. HSQC data was collected with 320 scans.

CD

Far-UV (202-250 nm) CD spectra were collected on a JASCO J-810 CD spectropolarimeter with a quartz cell having 1 cm path length. The concentration of INGAPrP was kept constant at 25 μg/mL for all CD measurements. An initial spectrum was collected in 5 mM Tris, 30 mM NaCl at pH 8.5 at 22.0°C. This sample was then heated to 90.0°C by a Peltier temperature control and another spectrum collected after the

protein was allowed to denature for 15 minutes. Finally, CD data was collected in the presence of 20%, 30%, 40%, 50%, and 60% 2,2,2-Trifluoroethanol (TFE) at 22.0°C. Each spectrum was compiled by averaging four scans and all spectra were baseline-corrected by subtracting out the buffer spectrum.

REFERENCES

Akerblom, H.K., Vaarala, O., Hyoty, H., Ilonen, J, Knip, M. (2002). Environmental factors in the etiology of type 1 diabetes. *Amer J Medical Genetics* **115**, 18-29.

Bailes, B.K. (2002). Diabetes mellitus and its chronic complications. *AORN J* **76**, 266-282.

Bouwens, L. (1998). Transdifferentiation versus stem cell hypothesis for the regeneration of islet beta-cells in the pancreas. *Microscopy Res Technique* **43**, 332-336.

Brand, S.J., Tagerud, S., Lambert, P., Magil, S.G., Tatarkiewicz, K., Doiron, K., Yan, Y. (2002). Pharmacological treatment of chronic diabetes by stimulating pancreatic beta-cell regeneration with systemic co-administration of EGF and gastrin. *Pharmacol Toxicol* **91**, 414-420.

Bukhtiyarova, M., Northrop, K., Chai, X., Casper, D., Karpusas, M., Springman, E. (2004). Improved expression, purification, and crystallization of p38alpha MAP kinase. *Protein Expr Purif* **37**, 154-161.

Buteau, J., Foisy, S., Joly, E., Prentki, M. (2003). Glucagon-like peptide 1 induces pancreatic beta-cell proliferation via transactivation of the epidermal growth factor receptor. *Diabetes* **52**, 124-132.

Cantenys, D., Portha, B. et al. (1981). Histogenesis of the endocrine pancreas in newborn rats after destruction by streptozotocin. *Virchows Arch. (Cell Pathol)* **35**, 109-122.

Clark E. (1998). Refolding of recombinant proteins. *Curr Opin Biotech* **9**, 157-163.

Cotterell, A.A. and Kenyon, N.S. (2002). Alternatives to immunosuppressive drugs in human islet transplantation. *Curr Diab Rep* **2**, 377-382.

Diallo, E.M., Thompson, D.L., Koenig, R.J. (2005). A method for efficient production of recombinant thyroid hormone receptors reveals that receptor homodimer-DNA binding is enhanced by the coactivator TIF2. *Protein Expr Purif* **40**, 292-298.

Ding, V.D.H., et al. (2002). Regulation of insulin signal transduction pathway by a small-molecule insulin receptor activator. *Biochem J* **367**, 301-306.

Dor, Y., Brown, J., Martinez, O.I., Melton D.A. (2004). Adult pancreatic □-cells are formed by self-duplication rather than stem-cell differentiation. *Nature* **429**, 41-46.

Fenjves, E.S., Ricordi, C. (2000). Gene therapy for type 1 diabetes. *Exp Opin Ther Patents* **10**, 325-331.

Flores, L.E., Garcia, M.E., Borelli, M.I., Zotto, H.D., Alzugaray, M.E., Maiztegui, B., Gagliardino, J.J. (2003). Expression of islet neogenesis-associated protein in islets of normal hamsters. *J Endocrinology* **177**, 243-248.

Gagliardino, J.J., Del Zotto, H., Massa, L., Flores, L.E., Borelli, M.I. (2003). Pancreatic duodenal homeobox-1 and islet neogenesis-associated protein: a possible combined marker of activateable pancreatic cell precursors. *J Endocrinology* **177**, 249-259.

George, M., Ayuso, E., Casellas, A. et al. (2002). □ cell expression of IGF-I leads to recovery from type 1 diabetes. *J Clin Invest* **109**, 1153-1163.

Gold, G., Broderick, C., Carfanga, M., Pittenger, G., Rafaeloff, R., Reifel-Miller, A., Borts, T., Hale, J., Chargay, L. (1998). INGAP treatment improves glycemic control in SZ diabetic hamsters. *Diabetes* **47** (Suppl. 1), A253.

Gray, D.W. (1997). Encapsulated islet cells: the role of direct and indirect presentation and the relevance to xenotransplantation and autoimmune recurrence. *Br Med Bull* **53**, 777-788.

Gu, D., Sarvetnick, N. (1993). Epithelial cell proliferation and islet neogenesis in IFN-gamma transgenic mice. *Development* **118**, 33-46.

Gu, Z., Su, Z., Janson, J. (2001). Urea gradient size-exclusion chromatography enhanced the yield of lysozyme refolding. *J Chromatography A* **918**, 311-318.

Heber-Katz, E. (2004). "Regeneration: stem cells and beyond." Springer, New York City.

Hedayati M.A., Grove, D.E., Steffen, S.E., Bryant, F.R. (2005). Expression and purification of the SsbB protein form Streptococcus pneumoniae. *Protein Expr Purif* **43**, 133-139.

Hughes, H.J. (1947). Cyclical changes in the islets of Langerhans in the rat pancreas. *J Anat* **81**, 82-92.

Hui, H., Wright, C., Perfetti, R. (2001). Glucagon-like peptide 1 induces differentiation of islet duodenal homeobox-1-positive pancreatic ductal cells into insulin-secreting cells. *Diabetes* **50**, 785-796.

Keymeulen, B., Vandemeulebroucke, E., Ziegler, A.G., et al. (2005). Insulin needs after CD3-antibody therapy in new-onset type 1 diabetes. *N Eng J Med* **352**, 2598-2608.

Kobayashi, S., Akiyama, T., Nata, K. et al. (2000). Identification of a receptor for Reg (Regenerating gene) protein, a pancreatic β-cell regeneration factor. *J Biol Chem* **275**, 10723-10726.

Kolodka, T.M., Finegold, M., Moss, L., Woo, S.L.C. (1995). Gene therapy for diabetes mellitus in rats by hepatic expression of insulin. *Proc Natl Acad Sci USA* **92**, 3293-3297.

Kulis, M.D. and Shuker, S.B. (2004). β-cell regeneration: a potential cure for type 1 diabetes. *Expert Opin Ther Patents* **14**, 599-605.

Lernmark, A. (1999). Type 1 diabetes. *Clinical Chem* **45**, 1331-1338.

Liang, P. and Pardee, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967-971.

Maria-Engler, S.S., Mares-Guia, M., Correa, M.L. et al. (2001). Microencapsulation and tissue engineering as an alternative treatment of diabetes. *Braz J Med Biol Res* **2001**, 691-697.

Mashima, H., Yamada, S., Tajima, T., Seno, M., Yamada, H., Takeda, J., Kojima, I. (1999). Genes expressed during the differentiation of pancreatic AR42J cells into insulin-secreting cells. *Diabetes* **48**, 304-309.

Mathis, D., Vence, L., Benoist, C. (2001). β-cell death during progression to diabetes. *Nature* **414**, 792-798.

Moller, D.E. (2001). New drug targets for type 2 diabetes and the metabolic syndrome. *Nature* **414**, 821-827.

Moustakas, A.K. and Papadopoulos, G.K. (2002). Molecular properties of HLA-DQ alleles conferring susceptibility to or protection from insulin-dependent diabetes mellitus: keys to the fate of islet β-cells. *Amer J Medical Genetics* **115**, 37-47.

Nielsen, J.H., Svensson, C., Galsgaard, E.D., Moldrup, A., Billestrup, N. (1999). Beta cell proliferation and growth factors. *J Mol Med* **77**, 62-66.

Notkins, A.L. (2002). Immunologic and genetic factors in type 1 diabetes. *J Biol Chem* **277**, 43545-43548.

Okamoto, H. (1999). The Reg gene family and Reg proteins: with special attention to the regeneration of pancreatic β-cells. *J Hepatobiliary Pancreat Surg* **6**, 254-262.

Pittenger, G.L., Rosenberg, L., Vinik, A.I. (1991). Partial purification and characterization of ilotropin, a pancreatic-islet specific growth factor. *J Cell Biol* **115**, 270A.

Probiodrug (2002). Patent US6500804.

Qureshi, S.A. et al. (2000). Activation of insulin signal transduction pathway and anti-diabetic activity of small molecule insulin receptor activators. *J Biol Chem* **275**, 36590-36595.

Rafaeloff, R., Qin, X.F., Barlow, S.W., Rosenberg, L., Vinik, A.I. (1996). Identification of differentially expressed genes induced in pancreatic islet neogenesis. *FEBS Letts*, **378**, 219-223.

Rafaeloff, R., Pittenger, G.L., Barlow, S.W., Qin, X.F., Yan, B., Rosenberg, L., Duguid, W.P., Vinik, A.I. (1997). Cloning and sequencing of the pancreatic islet neogenesis associated protein (INGAP) gene and its expression in islet neogenesis in hamsters. *J Clin Invest* **99**, 2100-2109.

Risbud, M.V. and Bhonde, R.R. (2002). Models of pancreatic regeneration in diabetes. *Diabetes Res Clin Practice* **58**, 155-165.

Rosenberg, L. (1995). *In vivo* cell transformation: neogenesis of beta cells from pancreatic ductal cells. *Cell Transplantation* **4**, 371-383.

Rosenberg, L., Rafaeloff, R., Class, D., Kakugawa, Y., Pittenger, G.L., Vinik, A.I., Duguid, W.P. (1995). Induction of islet cell differentiation in the hamster—further support for a ductal origin. *Pancreas* **13**, 38-46.

Rosenberg, L. (1998). Induction of islet cell neogenesis in the adult pancreas: the partial duct obstruction model. *Microscopy Research Technique* **43**, 337-346.

Rosenberg, L., Wang, R.N., Li, J.M., Pittenger, G., Duguid, W., Vinik, A. (2000). INGAP peptide increases β-cell mass and insulin content in adult hamsters and reverses STZ-diabetes. *Diabetes* (Suppl) **1**, A256.

Saltiel, A.R. and Kahn, C.R. (2001). Insulin signaling and the regulation of glucose and lipid metabolism. *Nature* **414**, 799-806.

Sandgren, E.P., Luetteke, N.C., Palmiter, R.D., Brinster, R.L., Lee, D.C. (1990). Overexpression of TGFβ in transgenic mice: induction of epithelial hyperplasia, pancreatic metaplasia, and carcinoma of the breast. *Cell* **61**, 1121-1135.

Sasahara, K., Yamaoka, T., Moritani, M., Yoshimoto, K., Kuroda, Y., Itakura, M. (2000). Molecular cloning and tissue-specific expression of a new member of the regenerating protein family, islet neogenesis-associated protein-related protein. *Biochimica Biophysica Acta* **1500**, 142-146.

Schwartz, S. and Bornfeldt, K. (2003). How does diabetes accelerate atherosclerotic plaque rupture and arterial occlusion? *Front Biosci* **8**, 1371-1383.

Shapiro, A.M.J., Lakey, J.R.T., Ryan, E.A., Korbutt, G.S., Toth, E., Warnock, G.L., Kneteman, N.M., Rajotte, R.V. (2000). Islet transplantation in seven patients with type 1 diabetes mellitus using a glucocorticoid-free immunosuppressive regimen. *N Eng J Med* **343**, 230-238.

Shaw, J.W. and Latimer, E.O. (1926). Regeneration of pancreatic tissue from the transplanted pancreatic duct in the dog. *Am J Physiol* **76**, 49-53.

Shi, Y., Jiang, C., Chen, Q., Tang, H. (2003). One-step on-column affinity refolding purification and functional analysis of recombinant human VDAC1. *Biochem Biophys Res Comm* **303**, 475-482.

Shuker, S.B., Hajduk, P.J., Meadows, R.P., Fesik, S.W. (1996). Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531-1534.

Soo Kim, Y., Lee, J., Shin, J., Kim, H., Kim, C. (2003). Enhancement of mouse pancreatic regeneration and HIT-T15 cell proliferation with rat pancreatic extract. *Biochem Biophys Res Comm* **309**, 528-532.

Soria, B. (2001). *In vitro* differentiation of pancreatic □-cells. *Differentiation* **68**, 205.

Sreerama, N., Woody, R.W. (1993). A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal Biochem* **209**, 32-44.

Stryer, L. (1995). "Biochemistry." 4th ed. W.H. Freeman and Company, New York City.

Sumi, S. and Tamura, K. (2000). Frontiers of pancreas regeneration. *J Hepatobiliary Pancreat Surg* **7**, 286-294.

Takatori, A, Ohta, E., Toshiaki, I., Horiuchi, K., Ishii, Y., Itagaik, S., Kyuwa, S., Yoshikawa, Y. (2003). Protective effects of probucol treatment on pancreatic □-cell function of SZ-induced diabetic APA hamsters. *Exp Anim* **52**, 317-327.

Terazono, K., Yamamoto, H., Takasawa, S., Shiga, K., Yonemura, Y., Tochino, Y., Okamoto, H. (1988). A novel gene activated in regenerating islets. *J Biol Chem* **263**, 2111-2114.

The Merck Manual of Medical Information (1997). Simon & Schuster Inc., New York City.

The pET System Manual (1997). Novagen. 7th ed.

The QIAexpressionist (2001). A handbook for high-level expression and purification of 6xHis-tagged proteins. 5th ed.

Thule, P.M., Liu, J.M. (2000). Regulated hepatic insulin gene therapy of STZ-diabetic rats. *Gene Therapy* **7**, 1744-1752.

Tuch, B.E., et al. (1998). Transplantation of genetically engineered insulin-producing hepatocytes into immunoincompetent mice. *Transplant Proc* **30**, 473.

U.S. Department of Health and Human Services (2000). Patent WO0009666.

Verge, C.F., Stenger, D., Bonifacio, E., Colman, P.G., Pilcher, C., Lingley, P.J., Eisenbarth, G.S. (1998). Combined use of autoantibodies (IA-2) autoantibody, GAD autoantibody, insulin autoantibody, cytoplasmic islet cell antibodies in type 1 diabetes: combinatorial islet autoantibody workshop. *Diabetes* **47**, 1857-1866.

Vetere, A., Marsich, E., Di Piazza, M., Koncan, R., Micali, F., Paoletti, S. (2003). Neurogenin-3 triggers β-cell differentiation of retinoic acid-derived endoderm cells. *Biochem J* **371**, 831-841.

Vinik, A.I., Pittenger, G.L., Rafaeloff-Phail, R., Barlow, S.W. (1998). "High level of expression of INGAP in bacterial and eukaryotic cells." Patent 5804421.

Watada, H., Kajimoto, Y., Miyagawa, J., et al. (1996). PDX-1 induces insulin and glucokinase gene expression in βTC1 clone 6 cells in the presence of betacellulin. *Diabetes* **45**, 1826-1831.

Watanabe, T., Yonemura, Y., Yonekura, H., Suzuki, Y., et al. (1994). Pancreatic beat-cell replication and amelioration of surgical diabetes by Reg protein. *Proc Natl Acad Sci USA* **91**, 3589-3592.

Wilkin, T.J. (1998). Neogenesis of islet cells. *Diabetes Metab Rev* **14**, 329-335.

Yamaoka, T., Itakura, M. (1999). Development of pancreatic islets. *Int J Mol Med* **3**, 247-261.

Yamaoka T. (2001). Gene therapy for diabetes mellitus. *Curr Mol Med* **1**, 325-337.

Yamaoka, T. (2002) Regeneration therapy of pancreatic β cells: towards a cure for diabetes? *Biochem Biophys Res Comm* **296**, 1039-1043.

Yamaoka, T. (2003). Regeneration therapy for diabetes mellitus. *Expert Opin Biol Ther* **3**, 425-433.

Yoon, J., Jun, H. (2002). Recent advances in insulin gene therapy for type 1 diabetes. *TRENDS in Mol Med* **8**, 62-68.

Zhang, S., Zubay, G., Goldman, E. (1991). Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates. *Gene* **105**, 61-72.

Zimmet, P., Alberti, K.G.M.M., Shaw, J. (2001). Global and societal implications of the diabetes epidemic. *Nature* **414**, 782-787.

VITA

Michael Kulis, Jr. was born in Syracuse, NY to Nancy and Michael Kulis on October 7, 1978. He attended Moravia High School in the central New York area, where he graduated as valedictorian of the class of 1996. Michael then went on to Cornell University in Ithaca, NY. He studied biology as an undergraduate and became interested in drug development using biochemical approaches. Upon graduation from Cornell in 2000, Mike moved to Atlanta, GA to study biochemistry in the graduate program at Georgia Institute of Technology. Upon completion of his doctoral work at Georgia Tech, Mike will be doing postdoctoral research with Dr. Wesley Burks at Duke University. At Duke, Mike anticipates doing high-impact work in the field of food allergies, particularly peanut allergy.